

**Visualizing dynamics of the central dogma in living single cells**

by

**Yue Guo**

B.Sc., Nanjing University, 2012

M.Sc., Institute of High Energy Physics, Chinese Academy of Sciences, 2015

Submitted to the Graduate Faculty of the  
Dietrich School of Arts and Sciences in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy

University of Pittsburgh

2022

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

**Yue Guo**

It was defended on

July 21, 2022

and approved by

Dr. Eric Swanson, Professor, Department of Physics and Astronomy

Dr. Yan Xu, Professor, Department of Physics and Astronomy

Dr. Hanna Salman, Associate Professor, Department of Physics and Astronomy

Dissertation Director: Dr. Robin E. C. Lee, Associate Professor, Department of Computational  
and Systems Biology

Dr. Xiao-Lun Wu, Professor, Department of Physics and Astronomy

Copyright © by Yue Guo

2022

## **Visualizing dynamics of the central dogma in living single cells**

Yue Guo, PhD

University of Pittsburgh, 2022

Imaging the central dogma at single molecule resolution in single cells reveals the spatial and temporal dynamics and ubiquitous cell-to-cell variability of molecular events. In my dissertation work, I first implement SunTag-based translation reporter and quantify the copy numbers of signaling molecule NEMO in NF- $\kappa$ B pathway to illustrate the use of state-of-the-art single molecule approach to study the central dogma. Transcription process to produce mRNAs as the start of the central dogma is the chief regulator of gene expression. However, most approaches to image mRNAs require cell fixation or have limited single-molecule sensitivity for live-cell applications. I therefore develop and characterize SunRISER, an approach for long-term imaging of mRNA in living cells. SunRISER employs SunTag as a scaffold to achieve fluorescence signal amplification of coat proteins and enhance contrast of mRNAs. Although the naïve design is impractical, with inconsistent fluorescent properties that complicate mRNA detection, I optimize the approach using computational and synthetic biology to achieve robust and unambiguous detection of individual mRNAs. SunRISER-labeled mRNAs are resistant to photobleaching and the design principle is generalizable for robust whole-cell mRNA imaging experiments with orthogonal tagging systems. SunRISER variants using shorter 8x and 10x stem-loop arrays (SunRISER SRv1.1 and SRv1.2, respectively) also result in consistent mRNA labeling and detection, while reducing the size of alterations to target mRNA sequences. As an application of SunRISER, I interrogate mitotic inheritance of mRNA molecules during a variety of stresses. When observed over the period of cell doubling time, it is found that mitotic mRNA inheritance is equally partitioned in standard growth conditions and that inflammatory stress or nutrient

limitation can enhance diversity among post-mitotic sister cells. SunRISER can be applied to other RNA species with further modification. Taken together, SunRISER enables a window into living cells to observe aspects of the central dogma in addition to roles of mRNAs in rare and dynamical trafficking events.

## Table of Contents

Acknowledgements .....	xii
<b>1.0 Chapter 1. Introduction and scope.....</b>	<b>1</b>
<b>1.1 Single cell measurement provides insight into heterogeneity and dynamical properties .....</b>	<b>1</b>
<b>1.2 mRNA production is a pivotal process in the central dogma .....</b>	<b>4</b>
<b>1.3 Techniques to visualize single mRNA molecules in fixed cell.....</b>	<b>6</b>
<b>1.4 Live cell imaging of single mRNAs has limited capability for long-term imaging... </b>	<b>8</b>
<b>1.5 Scope of this dissertation.....</b>	<b>9</b>
<b>2.0 Chapter 2 Antibody-epitope pairs in visualizing gene expression dynamics .....</b>	<b>10</b>
<b>2.1 Signal amplification principle with antibody-epitope pairs .....</b>	<b>11</b>
<b>2.2 Real-time imaging of single RNA translation dynamics in living cells.....</b>	<b>14</b>
<b>2.3 dNEMO: a tool for quantifying fluorescent puncta .....</b>	<b>18</b>
<b>2.4 Application of dNEMO: Estimating signaling molecule numbers in signaling complex.....</b>	<b>27</b>
<b>2.5 Naïve Design of SunRISER --- Use of SunTag for labeling of single mRNAs .....</b>	<b>31</b>
<b>2.6 Concluding remarks .....</b>	<b>35</b>
<b>3.0 Chapter 3 Development and optimization of SunRISER.....</b>	<b>36</b>
<b>3.1 Modeling of SunRISER system to assist in optimization of SunRISER.....</b>	<b>37</b>
<b>3.2 Experimental validation of optimal expression ratio predicted by model .....</b>	<b>45</b>
<b>3.3 Refinement of SunRISER for whole-cell mRNA imaging .....</b>	<b>51</b>

3.4 SunRISER variants with small stem-loop arrays and reduced plasmid requirements .....	64
3.5 Concluding remarks.....	69
4.0 Chapter 4. Characterization and application of SunRISER .....	70
4.1 Characterization of SunRISER system .....	71
4.2 Application of SunRISER: Variability of mitotic mRNA inheritance between sister cells.....	77
4.3 Concluding remarks.....	83
5.0 Chapter 5 Materials and methods.....	84
5.1 Plasmid construction .....	85
5.2 Live cell imaging and quantification of single mRNA spots with dNEMO.....	86
5.3 Stress treatments .....	87
5.4 smFISH probes and image acquisition.....	87
5.5 Fixed-cell Immunofluorescence.....	88
5.6 Statistical analysis.....	89
5.7 Half-life and diffusion constant measurement.....	89
5.8 mRNA inheritance during mitosis analysis.....	90
6.0 Chapter 6 Discussion .....	91
6.1 Limitation of SunRISER.....	94
6.2 Future perspectives in application of SunRISER.....	95
Appendix A.....	97
Bibliography .....	122

## List of Tables

<b>Table 3-1 Summary of SunTag/PCP plasmid variants. ....</b>	<b>61</b>
<b>Table 3-2 Molecular weight for SunRISER variants and other mRNA labeling techniques for labeling single mRNAs.....</b>	<b>62</b>
<b>Table 3-3 Summary of parameters used in simulation.....</b>	<b>63</b>

## List of Figures

<b>Figure 2-1 The schematic of SunTag labeling.</b> .....	<b>13</b>
<b>Figure 2-2 The schematic of SunTag-based labeling of nascent peptides.</b> .....	<b>16</b>
<b>Figure 2-3 Visualization of translating polysomes in HEK293 cells.</b> .....	<b>17</b>
<b>Figure 2-4 The à trous wavelet transform on simulated and experimental images.</b> .....	<b>22</b>
<b>Figure 2-5 Puncta with axial overlap are individually resolved.</b> .....	<b>23</b>
<b>Figure 2-6 Local background information is used to correct intensity values measured for each punctum.</b> .....	<b>24</b>
<b>Figure 2-7 Identification of smFISH transcripts in fixed-cell images.</b> .....	<b>25</b>
<b>Figure 2-8 Quantification of translating polysomes with dNEMO.</b> .....	<b>26</b>
<b>Figure 2-9 Polysome calibration to estimate numbers of EGFP-NEMO molecules per complex.</b> .....	<b>30</b>
<b>Figure 2-10 Design of SunRISER, a SunTag based reporter for imaging signal-enriched mRNA.</b> .....	<b>33</b>
<b>Figure 2-11 Naïve design shows significant variability in spot intensity and size.</b> .....	<b>34</b>
<b>Figure 3-1 The schematic of signal and background definition when using SunRISER to label an mRNA.</b> .....	<b>41</b>
<b>Figure 3-2 Parameter sweeps in the space spanned by number of molecules for scFv-GFP and nxSunTag-PCP (left, (A) n=24 (B) n=5 (C) n=10) using a computational model to calculate number of GFP molecules per mRNA.</b> .....	<b>42</b>

<b>Figure 3-3 Heatmaps of mean number of GFP molecules per mRNA (left) and signal-to-background ratio (right) in the parameter space spanned by the number of molecules for scFv-GFP and ‘n’xSunTag-PCP (from left to right, 5x, 10x and 24x).....</b>	<b>44</b>
<b>Figure 3-4 Quantification of strengths for constitutive promoters in HeLa cells. ....</b>	<b>46</b>
<b>Figure 3-5 Varying promoters for scAB-GFP and 10xSunTag-PCP corroborates model predictions. ....</b>	<b>48</b>
<b>Figure 3-6The preliminary version of SunRISER labels only cytoplasmic mRNA because SunTag-PCP is excluded from nuclei.....</b>	<b>49</b>
<b>Figure 3-7 Modifications to SunTag-PCP alter sub-cellular distribution of SunRISER components. ....</b>	<b>53</b>
<b>Figure 3-8 The optimal SunRISER design allows long-term imaging of single mRNA molecules throughout the cell. ....</b>	<b>55</b>
<b>Figure 3-9 SunRISER-labeled mRNAs comparing weak and strong promoters in the detection plasmid, and expression in A549 cells. ....</b>	<b>57</b>
<b>Figure 3-10 Design principles of SunRISER are generalizable to orthogonal stem-loops and protein tagging systems. ....</b>	<b>58</b>
<b>Figure 3-11 Model calibrated to orthogonal RNA hairpin and coat protein systems are comparable with previous observations for PP7.....</b>	<b>59</b>
<b>Figure 3-12 Design of SunRISER variants. ....</b>	<b>66</b>
<b>Figure 3-13 Quantification and representative images of different configurations for SunRISER.....</b>	<b>67</b>
<b>Figure 4-1 Comparison between mRNAs labeled with SunRISER SRv1 and PP7-PCP-GFP. ....</b>	<b>72</b>

<b>Figure 4-2 SunRISER is resistant to photobleaching and robustly labels mRNA molecules during prolonged imaging conditions. ....</b>	<b>73</b>
<b>Figure 4-3 Characterization of translation, half-life, and cytoplasm-to-nucleus ratio for SunRISER-labeled mRNA. ....</b>	<b>75</b>
<b>Figure 4-4 Diffusion of SunRISER labeled mRNAs. ....</b>	<b>76</b>
<b>Figure 4-5 Long-term imaging to monitor mRNA partitioning during cell division. ....</b>	<b>79</b>
<b>Figure 4-6 Symmetry and asymmetry of mitotic mRNA partitioning identified by k-means clustering and active transcription during mitosis. ....</b>	<b>81</b>
<b>Figure 4-7 Mitotic mRNA partitioning between daughter cells during stress. ....</b>	<b>82</b>

## Acknowledgements

I would like to thank many people for this dissertation. First, I am deeply grateful to my supportive and creative supervisor Dr. Robin Lee for his guidance and inspiration throughout all my years of studying in his lab. When I first started my rotation in his lab, I had limited knowledge and skills in molecular and systems biology, but I was not only stricken by his quick thinking and fascinating research ideas but also by open-mindedness and belief to a student with physics background. I am grateful for the opportunity to learn from his as a role model of rigorous and innovative scientist and his expertise and dedication were something I can always count on and look up to.

I am grateful to all members of Lee lab. Specifically, I would like to thank Gabriel Kowalczyk for helping me with dNEMO analysis in the process of developing SunRISER, David Schipper for introducing me to the lab and being a great lab manager, Dr. Agustin Cruz for sharing broad knowledge and educational collaboration, Chaitanya Mokashi for patient advice and intriguing discussions, Steven Smeal for warm encouragement and helpful comments.

I would also like to thank Dr. Yi-jiun Chen and Dr. Qiuhong Zhang for their mentorship and valuable resource to kickstart my wet-lab journey.

I would like to thank my dissertation committee members Dr. Xiao-Lun Wu, Dr. Eric Swanson, Dr. Yan Xu and Dr. Hanna Salman for their time, wisdom and perspectives.

Last but not least, I would like to thank my family. There are no words enough to express how I miss them and how grateful I am to their everlasting love, trust and support. My choice of studying abroad is even harder to them than myself but they never questioned my decision. I want to dedicate this dissertation to my family.

## **1.0 Chapter 1. Introduction and scope**

### **1.1 Single cell measurement provides insight into heterogeneity and dynamical properties**

Single cells as the basic computing unit of organisms have been the major focus of biological studies in more than a century and become the fundamental resolution to understand multi-cellular processes. A central question to be answered in biology is how distinct individual cells contribute to define organism as a whole. How do cells differentiate from each other and make different cell fate decisions? How do different cells communicate to their neighbors and respond to their native environment coordinately? By choosing their dedicated fates and performing unique functions, single cells are specialized workers that operate the whole machine and require precise instructions. Failure to understand single cell behavior will impose difficulty in interpreting biological principles. One contributing factor to this difficulty that necessitates single cell analysis is the ubiquitous cell-to-cell heterogeneity in any subpopulation of “seemingly identical” cells [1, 2]. Genetic heterogeneity is naturally present in any population of cells due to random mutations. Non-genetic or phenotypic heterogeneity [3-8] is more prevalent and now recognized as one efficient way cells exploit to cope with various environmental cues. Non-genetic heterogeneity can be classified as extrinsic or intrinsic based on the origin [9]. Extrinsic heterogeneity refers to diverse cellular responses to particular physical and biological environments and stems from complex culture environment such as different neighboring cells or specific chemical gradients and is only reflected in subpopulations exposed to the triggering factors. Intrinsic heterogeneity conversely cannot attribute to obvious external source and arises from inherent diversifications within a complex system and often involves symmetry-breaking

distributing events that create discrete distinction or stochastic biochemical reactions resulting continuous differentiation. For instance, molecules present at low copy numbers are prone to intrinsic heterogeneity resulted from random production, degradation, and partition. Another confounding factor the requires single-cell resolution studies is the dynamical nature of every biological process [10-14]. Traditionally, biological principles including regulatory networks and signaling pathways are established via invasive biochemical assays that isolate target molecules to measure amounts and characterize interactomes, which barely preserves the native environment where the biological processes take place and let alone the spatiotemporal dynamics. But more and more experimental evidence shows that cells can respond distinctly to the same stimulus but with different temporal dynamics such as duration and oscillation frequency [15, 16] so average analysis from cell populations could misrepresent cells receiving distinct spatiotemporal dynamic activation patterns.

Single cell studies are enabled by blooming of single-cell resolution techniques [17-24]: flow cytometry, fluorescence microscopy and single-cell sequencing. Flow cytometry[25] allows hundreds of thousands of single cells to flow through a multiple-laser system and be recorded as individual data points in a multi-parametric space including size, viability and fluorescence properties that mark specific labeled molecules. Flow cytometry is a high-throughput technique and widely used to assess protein level and gene expression to classify cell types. However, it is impossible to follow single cells over time and retrieve spatial information with conventional flow cytometry. Single cell sequencing [26-30] has become the most powerful technique demonstrating comprehensive insights into single cell genomics [31], transcriptomics [32], and proteomics [33, 34] to provide fundamental and transformative understanding of many diseases including cancers [35] and COVID-19 [36]. The successful implementation of single cell sequencing technology

involves isolation of single cells, extraction and barcoding into sequencing library and final sequencing with commercial platforms. But chemical or physical manipulation of single cells can potentially cause unintended DNA damage or cell stress and therefore affect detection accuracy and introduce higher noise. Additionally, cells are no longer maintained at their native environment during isolation and the cell behavior and gene expression profile could also be perturbed unexpectedly. By contrast, fluorescence microscopy affords direct imaging of single cells without disturbing their innate surroundings. Live cell imaging is further equipped with the capability to retain cell identity and capture temporal complexity in biological processes and is the essential tool to probe dynamical cellular events. Although fluorescence microscopy methods suffer from low-throughput and sophisticated set-up but “seeing is believing” still motivates incessant improvement of instrumentation and labeling tools.

Eukaryotic cells establish specificity by interpreting a multitude of heterogeneous signals they encounter as inputs and computing appropriate responses. These signals often confer dynamics in duration and strength and co-present with other signals. Cells transmit signal information by activating multiple signaling pathways, usually manifested as protein location shift and protein complexes assembly. The core decoder of signaling pathway is transcription factors, which precisely control the expression of a specific set of target gene. Conventional biochemical measurements at fixed time points have depicted signaling networks and gene expression patterns and it is evident that multiple signaling pathways are interconnected where expressed proteins of target genes from one pathway could be inhibitors or activators for another pathway. This structure suggests that the temporal features of individual processes including duration and sequence are vital to the overall cellular responses and static portray of the signaling system is far from enough to define the whole story. Cells must be capable of translating external cues and encoding them as

various temporal patterns in signaling networks and also possess the mechanism to decode the dynamics and produce corresponding transcriptional responses [37, 38]. The non-linear nature has been revealed in many signaling pathways by live cell imaging and following cellular processes in the same cell over time [39]. For example, NF- $\kappa$ B signaling pathway can exhibit either oscillatory or sustained activation patterns when stimulated with different signals [40, 41] and therefore different gene expression profiles. P53 dynamics are also differently modulated in response to DNA damage caused by different radiation stimuli and doses and contribute to distinct cell fates [42, 43].

The behavior and fate of single cells are determined by complex cell signaling and gene expression regulation. Combined single-cell measurement with mathematical models, it is increasingly clear that heterogeneity and dynamics are universal features and governing principles in the control of cell fate. Single-cell studies also revealed new complexity and requires an even closer look of cellular processes at single molecule resolution.

## **1.2 mRNA production is a pivotal process in the central dogma**

The central dogma as the fundamental processes for genetic information flow from DNA to mRNA (transcription) and mRNA to protein (translation) is the master machinery where cells assimilate its innate state and environmental signals and generate appropriate responses for cell fitness and survival. The regulation of gene expression is the key step of single cell responding to internal and external stimuli as it would subtly adjust the mRNA and protein composition and drive cell fate decisions. The central dogma as a multistep process can be modulated at each step [44-47]. Starting from association of transcription factors with regulatory regions such as promoters

and enhancers, chromatin could undergo remodeling to an open state and then assembly of pre-initiation complex and recruitment of RNA polymerase II initiates the synthesis of mRNAs. After elongation, pre-mRNA is modified in nucleus for 5'-capping, splicing and 3'-end processing before nuclear export and mature mRNA is translated or degraded in cytoplasm [48-50]. This simplistic description leaves out all the complexity and molecular details for dynamic control of mRNA life-cycle by RNA binding-proteins but it is obvious that transcription upstream of translation is exposed to more stringent controls and can affect the final protein product to a larger extent as any regulation at mRNA level would be amplified and accumulated in downstream pathways.

“The central dogma of molecular biology”, DNA to RNA to protein, distilled by Francis Crick half a century ago [51] has been the dominant axis in understanding cell behaviors, but more and more studies have illustrated complications of the central dogma [52-54]. Reverse transcriptases [55, 56] read genetic information from RNA and convert into DNA form that can be integrated into genome, the process of which is prevalent in virus [57], prokaryotic [58] and eukaryotic cells [59] and enables addition or modification to genomic sequences. Reverse transcription has been widely applied to extract information from RNA transcripts in large scale, including RT-PCR and RNA-seq [60] and revealed gene expression profiles under particular stimuli. Some RNA molecules do not serve as templates to make proteins but undergo structural changes and possess catalytic functions [61]. Besides the distort of the classical information flow, information content has more inputs than upstream sequence codes. Posttranscriptional RNA processing [62, 63] and posttranslational protein modification [64] are discovered as widespread mechanisms to determine RNA and protein functional properties that cannot be gained from direct copy. DNA viewed as the most stable and inherent elements in the central dogma also subjects to

protein-based regulation [65, 66]. The twists of the central dogma have motivated closer studies of the key players: DNA, RNA and protein at unprecedented resolutions: single molecule at various time scales to recognize new conceptual formulations.

A growing body of evidence suggests that RNA is the most critical molecule to comprehensive and advanced understanding of the central dogma [67]. The birth and decay of mRNA molecules [68, 69] that contributes to the fluctuation in mRNA and protein copies is crucial for precise control of gene expression levels. Regulation of mRNA distributions at subcellular compartments [70, 71] is used to generate spatial patterns of gene expression. Continual progress has been made to understand transcription machinery, but the temporal details of molecular interactions are still elusive. All the mRNA processing steps can be easily misinterpreted when it is impossible to directly follow the events in real time. Therefore, it requires advances in imaging methodologies to understand gene expression by mRNA regulation, including transcriptional bursting, nuclear export, localized translation and decay mechanisms.

### **1.3 Techniques to visualize single mRNA molecules in fixed cell**

Intracellular localization of mRNA is known as an important mechanism to regulate localized protein production. In Situ Hybridization (ISH) has been invented to examine distribution of transcripts in fixed cell. Initially radioactively labeled probes were used for detection [72, 73] but these assays bear the disadvantage of low resolution and hazardous radiations. Later development of biotin as labeling molecules conjugated to specific antibody improves the applicability of ISH assays but still rely on long labeled probes [74]. In 1998, the Singer laboratory presented the first demonstration of fluorescence in situ hybridization achieving single molecule

resolution [75]. This early implementation of smFISH (single-molecule FISH) employs as few as five 50-nt-long oligonucleotides labeled with five fluorophores each, which makes nonspecific excessive probes indistinguishable from true target binding signals and therefore high false-positive rate. To improve signal-to-background, Raj et al. [76] proposed a refined version where a set of 48 20-nt-long probes each carrying single fluorophore is used and single transcript is detected with uniform intensity. Finally, with reliable detection efficiency smFISH has become the standard approach to label single mRNAs in fixed samples.

Measuring signaling dynamics and gene expression in the same single cell unravels the connection between signaling dynamics and quantitative change in gene expression. Lee et al. [77] quantified the dynamics of NF- $\kappa$ B with a fluorescent protein labeled RelA reporter and computed better predictors for NF- $\kappa$ B transcriptional response illustrated with end-point smFISH to be fold-change of nuclear NF- $\kappa$ B. They also proposed a mathematical model and discovered an incoherent feed-forward loop motif introduced by competition on NF- $\kappa$ B target promoters is capable of fold-change detection. smFISH studies also reveal the stochasticity of transcription process as it directly measures copy number of transcripts [78].

Despite the extensive spatial information and resolving power for single transcript, smFISH still lacks temporal dimension to follow mRNA expression dynamics in real time. As a snapshot approach, cells are fixed and permeabilized to chemically preserve cell states but are left with only the information from selected single time point, which might lead to misinterpretation of dynamics by inference. Rare and transient events are even difficult to capture with smFISH as it is impossible to predict the appropriate time point beforehand.

## 1.4 Live cell imaging of single mRNAs has limited capability for long-term imaging

mRNA molecules interact with RNA binding proteins throughout their lifespan to carry genetic information and provide precise spatiotemporal regulation within cells. Live-cell single molecule imaging techniques have enabled in-depth characterization of dynamics for mRNA processing steps, including transcription, translation, splicing, export, degradation, and interactions with ribonucleoprotein (RNP) granules [79-96]. However, continuous imaging of single mRNAs has numerous challenges coupled to low imaging sensitivity that is exacerbated by rapid photobleaching [97, 98]. Typically, live-cell detection of single mRNAs requires sophisticated imaging approaches and trade-offs that restrict spatial and temporal aspects of imaging experiments.

Bacteriophage-derived MS2 and PP7 stem-loops are extensively used for labeling mRNA molecules. In many applications, the reporter mRNA is tagged with 24x stem-loop copies in the 3' UTR and the corresponding coat protein (MCP or PCP) is fused with a fluorescent protein (FP). When co-expressed in the same cell, dimers of FP-fused coat proteins bind to each stem loop enabling visualization of mRNAs and active transcription sites by fluorescence microscopy [89, 98-102]. For most applications, coat proteins are also fused to a nuclear localization signal (NLS) to deplete unbound FP-MCP and FP-PCP from the cytoplasm, thereby increasing image contrast for mRNAs labeled in the cytoplasm [103-108]. Since the NLS on unbound FP-M/PCP will favor a nuclear localization, the reduced availability of free coat proteins will limit cycling of fluorescent proteins on cytoplasmic mRNAs. Consequently, depletion of fluorescence signals on cytoplasmic mRNA molecules resulting from dissociated and photobleached FP-M/PCP molecules is difficult to recover, which can constrain cytoplasmic mRNA detection over longer time scales. Although variants with increased stem-loop numbers enhance imaging sensitivity (up to 128 copies of

extended MS2 repeats; [109-111]), the bulky mRNA extension has the potential to perturb mRNA dynamics, and the reporter still suffers from photobleaching and limited resolution for single mRNAs.

### **1.5 Scope of this dissertation**

In this dissertation, I discuss the implementation of the state-of-the-art signal amplification SunTag technology to visualize and quantify aspect of central dogma and molecular signaling events in living cells. As a focus of my dissertation, I develop a novel live-cell mRNA reporter SunRISER that allows observation of mRNA at single molecule resolution for extended duration with resistance to photobleaching. Using SunRISER I investigate the kinetics of single mRNAs over the duration of one full day. By quantifying mRNA partitioning between sister cells during mitosis, I discover asymmetry mRNA inheritance in cell division upon stresses.

## 2.0 Chapter 2 Antibody-epitope pairs in visualizing gene expression dynamics

This chapter is adapted from:

JA Cruz\*, CS Mokashi\*, GJ Kowalczyk, **Y Guo**, Q Zhang, S Gupta, DL Schippner, REC Lee, A variable-gain stochastic pooling motif mediates information transfer from receptor assemblies into NF- $\kappa$ B, *Science Advances* 7 (30), eabi9410, 2021, of which I was was a co-author, with the following author contributions: Conceptualization: J.A.C., C.S.M., and R.E.C.L. Methodology: J.A.C., C.S.M., G.J.K., Y.G., S.G., D.L.S., Q.Z., S.W.S., and R.E.C.L. Software: C.S.M., G.J.K., S.G., and S.W.S. Formal analysis: J.A.C., C.S.M., G.J.K., Y.G., and D.L.S. Investigation: J.A.C., C.S.M., Y.G., C.S.M., and Q.Z. Writing—original draft: J.A.C., C.S.M., and R.E.C.L. Writing—review and editing: J.A.C., C.S.M., G.J.K., Y.G., S.A.G., D.L.S., Q.Z., and R.E.C.L. Visualization: J.A.C., C.S.M., and R.E.C.L. Supervision and funding acquisition: R.E.C.L.

GJ Kowalczyk, JA Cruz, **Y Guo**, Q Zhang, N Sauerwald, REC Lee, dNEMO: a tool for quantification of mRNA and punctate structures in time-lapse images of single cells, *Bioinformatics* 37 (5), 677-683, 2021, of which I was a co-author, with the following author contributions: Conceptualization, R.E.C.L.; Methodology, R.E.C.L., G.J.K., and N.S.; Software G.J.K., N.S., and R.E.C.L.; Image Acquisition, J.A.C., Y.G., Q.Z., and R.E.C.L.; Software Testing, J.A.C., Y.G., Q.Z.; Writing – Original Draft, G.J.K. and R.E.C.L.; Writing – Review & Editing, R.E.C.L., G.J.K., J.A.C., Y.G., and N.S.; Visualization, R.E.C.L. and G.J.K.; Funding Acquisition, R.E.C.L.; Supervision, R.E.C.L.

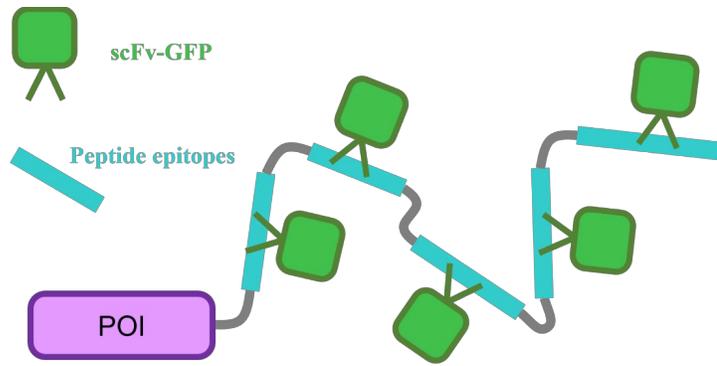
## 2.1 Signal amplification principle with antibody-epitope pairs

Varying copy numbers of functional sequences or proteins associated with a target substrate is a common regulatory strategy in biological systems. For example, promoter region consisting of several binding sites to one transcription factor (TF) facilitates recruiting multiple copies of TF to enhance activation and the presence of competing target sites for the same TF can reduce the effective copies of available TF [112]. Nuclear import of large proteins is mediated by the interaction of short peptide motifs known as nuclear localization signals (NLS) with transport machinery importin [113, 114]. Many proteins that are required to reside in nucleoplasm and perform specific functions in nucleus contain multiple NLS to prevent mislocalization with redundant nuclear import pathways [115, 116].

The principle of multimerization has been exploited for signal amplification in biological imaging applications. As described in last chapter, bacteriophage-derived stem-loops and complementary coat proteins are employed to introduce multiple copies of fluorescent protein to single mRNAs for visualization. The most common technique to detect single mRNA molecules smFISH [76, 117] also introduces a group of short fluorescence labeled DNA probes to target mRNA and enables uniform and accurate detection by eliminating noise from nonspecific binding. Similarly, antibody-epitope motifs have been developed to serve as protein recruitment platforms for bright labeling of proteins. SunTag [118] system is the most widely used genetically-encode antibody-epitope for fluorescent tagging. An scFv derived from a GCN4 antibody is labeled with superfolder GFP and 19-aa repeated peptides (V4) are linked by length optimized spacer to allow epitopes in the SunTag to be fully occupied by scFvs, generating ultra bright fluorescence signals (Figure 2-1). MoonTag [119] is another orthogonal antibody-peptide system to SunTag, where 15-

aa peptide is from the HIV envelope protein complex subunit gp41 and antibody is a 123-aa llama nanobody that binds the peptide in vitro with an affinity of 30 nM.

The antibody-epitope pairs including SunTag and MoonTag possess several advantages over direct fluorescent protein fusion for protein labeling: First, with extreme high signal amplification, proteins can be imaged with much lower illumination power and therefore subject to less phototoxicity for long-term imaging applications; Second, SunTag/MoonTag enables imaging proteins expressed at low levels without potential disturbance and artifacts introduced by overexpression; Third, by the time of visualizing tagged proteins, a pool of matured fluorescent protein is already formed and does not require further waiting time for protein folding, which provides an accessible approach to observe quick biological processes. The capability of immediate detection upon antibody binding with multiple pre-existing fluorescent proteins renders SunTag the potential to shed light on the core biological processes with super resolution.



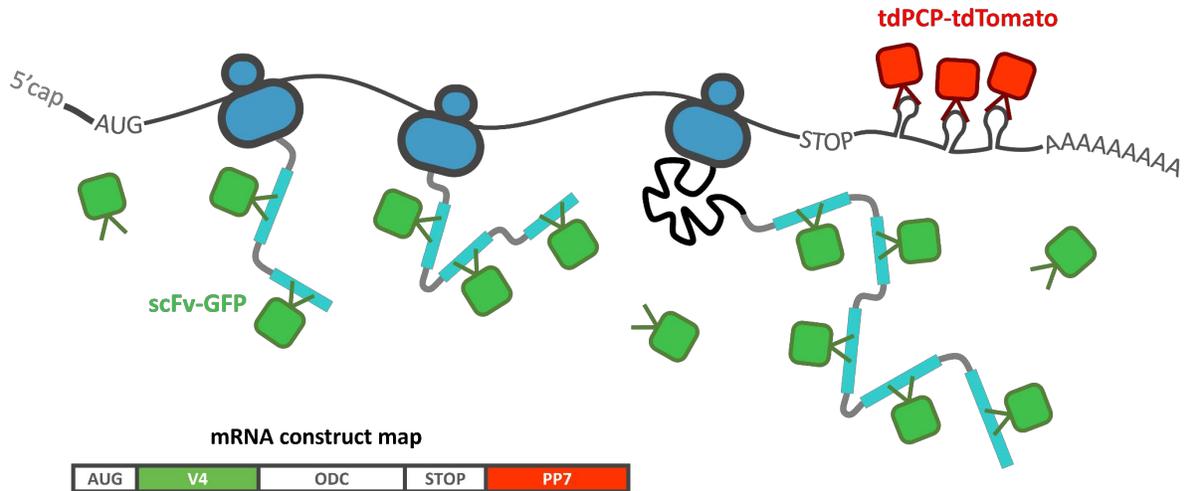
**Figure 2-1 The schematic of SunTag labeling.**

Short peptide sequences (turquoise) not native to mammalian cells bind to soluble GFP fused single chain variable fragment (scFv) antibodies (green) and can be used to label protein of interest (POI; purple) when multiple copies are attached with super high brightness. Illustration is redrawn based on concepts in [118].

## 2.2 Real-time imaging of single RNA translation dynamics in living cells

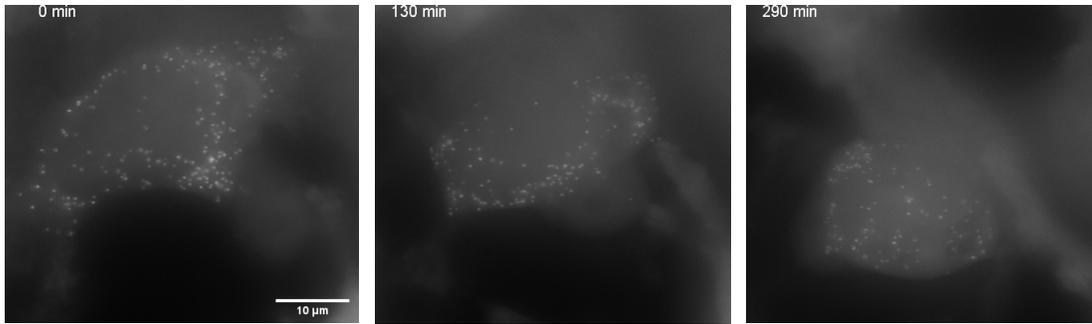
Translation as the final step for genetic information transfer from mRNA to protein is tightly regulated in space and time by a variety of mechanisms[120-124]. The regulation happens at all three stages of translation process: initiation, elongation and termination. Global regulation of protein synthesis is often exerted by changing activities and availabilities of translation initiation factors such as cap-binding protein eIF4E [125] and mRNA-specific regulation is dependent on particular motifs present in untranslated regions that alter recruitment of translation machinery [126, 127]. Ribosome profiling has been applied to study genome-wide translation dynamics as it maps the instantaneous position of ribosomes on almost all cellular mRNAs and provides in vivo quantitative measurement of protein synthesis rate [128-131]. However, this method lacks temporal resolution and assumes a uniform elongation rate among different mRNAs. Dissecting sophisticated mechanisms regulating translation rate and heterogeneity between mRNAs still requires highly accurate kinetic analysis at single molecule level. Single cell imaging approaches have difficulty in resolving nascent peptide chains at translation sites due to the relatively long maturation time with fluorescent proteins and are usually limited to measurement of average protein synthesis rate and inference of translation sites [83, 132, 133]. With the advent of SunTag principle, several groups have proposed use of antibody-epitope pairs to directly monitor the translation activity on individual mRNA molecules [93, 96, 134-136]. We adapted the approach from [93] to visualize translation of single mRNAs in HEK293 cells. Specifically, a reporter transcript encoding an array of 24xSunTag peptides followed by an ornithine decarboxylase (ODC) is co-expressed with scFv-GFP. ODC as one of the most short-lived proteins would promote degradation of ODC fused protein and help distinguish translating polysomes from translation products. 24 copies of PP7 stem-loop sequence are placed into the 3' UTR of target

mRNA and binding to tdTomato-fused PP7 coat proteins (tdTomato-tdPCP) allows independent readout of mRNA locations. During translation, single mRNA is occupied by a group of consecutive ribosomes called polysome and once the nascent SunTag peptides are translated from ribosomes, they will bind multiple copies of already fluorescent scFv-GFP and become detectable as bright GFP foci (Figure 2-2) in the cytoplasm as translating polysomes diffuse much slower than released protein products. In successful implementation of SunTag-based single molecule translation reporter, advanced microscopy such as confocal microscope is utilized to further improve signal-to-background ratio and the reporter system is stably integrated into target cell lines for better reproducibility. To validate the reporter system in different cell lines and test the applicability with standard microscope, we request the plasmids encoding the translation reporter and deliver them to HEK293 cells via transient transfection. 12 hours after transfection, we observed bright GFP spots in cytoplasm.



**Figure 2-2** The schematic of SunTag-based labeling of nascent peptides.

mRNA construct map depicts essential components for SunTag-based translation reporter. Reporter mRNA (black) is translated by multiple ribosomes (blue) and newly made V4 peptide (turquoise) is bound by scFv-GFP (green) and the whole translating complex will be seen in green fluorescence channel. tdPCP-tdTomato is recruited to PP7 stem-loops in 3'-UTR for visualization of individual mRNAs in red fluorescence channel. Illustration is redrawn based on concepts in [93].



**Figure 2-3 Visualization of translating polysomes in HEK293 cells.**

Time course for maximum intensity projection images of a representative HEK293 cell expressing SunTag-based translation reporter where individual translation polysomes appear as diffraction-limited spots. Cells were imaged 12hrs after transfection for 6 hr with a 10-minute frame rate.

Scale bar: 10  $\mu\text{m}$ .

### 2.3 dNEMO: a tool for quantifying fluorescent puncta

As discussed in previous sections, processes of the central dogma, such as single mRNA transcripts and sites of active protein translation and other regulatory multi-protein assemblies can be observed by fluorescence microscopy as punctate structures within the cell. Accurate detection and quantification of biological puncta is necessary to examine their roles in regulating cellular behaviors, and computational analysis is crucial for information extraction and often the rate-limiting step of experimental pipelines. Here, we developed detecting-NEMO (dNEMO), a free application that uses wavelet-based spot detection and supervised segmentation to detect and measure properties of fluorescent puncta in fixed-cell and time-lapse images. In contrast with many applications that completely automate analysis, dNEMO encourages users to interact directly with their data. Intuitive tools for cell segmentation, spot inspection and background correction, in addition to manual and automated selection of puncta based on quantifiable features (e.g. size, location and fluorescence) ensure that single-cell data are of the highest quality.

Wavelet-based approaches are used in image analysis for de-noising, compression and feature extraction with low computational cost [137, 138]. In wavelet-based feature extraction applications, the source image is decomposed into wavelet maps, a series of images where contrast is enhanced for particular spatial features. Since the wavelet transform sequentially applies a different convolution matrix at successive levels of the algorithm, the size and qualities of spatial features that are enhanced in each wavelet map can be modulated. To evaluate the effect of wavelet transformation, we used dNEMO to analyze three types of simulated and experimental images that contain diffraction-limited puncta (Figure 2-4). As the wavelet map transform level increases, zeros are progressively inserted into the  $\grave{a}$  trous wavelet convolution matrix. Comparing experimental images of diffraction-limited spots at the first level of the wavelet algorithm (L1

wavelet map), noise and the smallest puncta in the source image were enhanced. Consistent with previous findings [139, 140], the L2 wavelet map enhanced contrast for puncta at or near the diffraction limit. At higher levels, larger puncta were more resolved at the expense of reduced clarity for smaller puncta. Although users of dNEMO can select a wavelet map appropriate for their application, the L2 wavelet map effectively enhanced contrast in our three test images and was used to detect small molecular assemblies in subsequent experiments.

To identify fluorescent puncta near the diffraction limit, watershed and thresholding algorithms in dNEMO were used to segment wavelet maps for three types of imaging data. Puncta identified by watershed were then evaluated to prevent over-segmentation, where a single punctum with a noisy spatial distribution of fluorescence is erroneously segmented by the watershed into two or more puncta. To further prevent misidentification of a single punctum in consecutive slices of a 3D image, centroids in adjacent slices of an image stack are merged if their x-y coordinates are nearly identical. Using these approaches, the z-coordinate of each centroid can be calculated and we observed that distinct yet axially overlapping spots in a 3D image can be resolved by dNEMO (Figure 2-5).

Slow-varying background from non-specific dye accumulation and free fluorescent proteins that are not part of molecular assemblies, among other sources, will contribute to the measured intensity of a punctum. To correct for these effects, dNEMO collects local background pixel information from an annular ring that surrounds each punctum in the source image (Figure 2-6). Instead of using procedural generation of annuli for each punctum, the method in dNEMO operates directly on the wavelet map and consequently background correction is rapid.

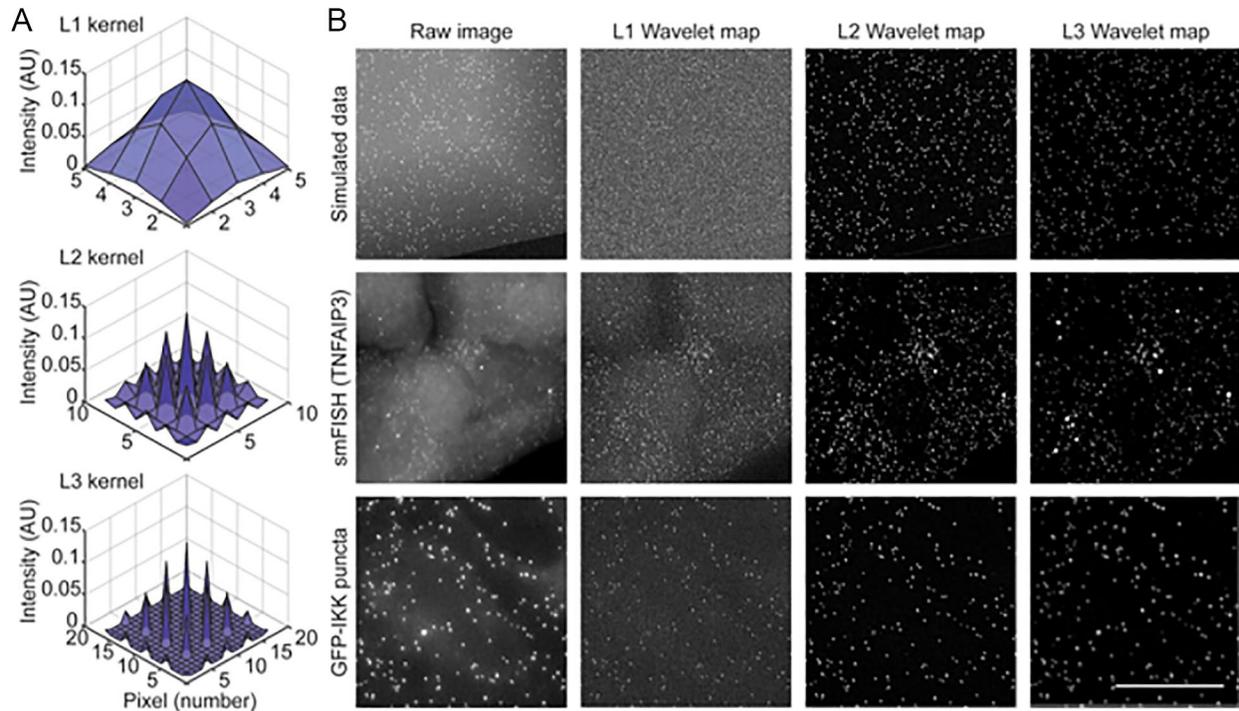
To associate and compare puncta between single cells, dNEMO contains an interactive polygon tool for manual cell segmentation (Figure 2-7). Cell segmentation paths can also be

imported into dNEMO for compatibility with results from other software. All puncta contained within the polygon are associated and puncta features, such as their number and distributions of fluorescent intensities among others, can be collated for each single cell. As a demonstration, we used dNEMO to detect single molecules of mRNA from smFISH images of TNF-induced expression of the NFKBIA gene using a previously collected dataset [77]. Although there was significant variability in transcript numbers when compared between single cells, the size and fluorescence intensity distributions of puncta were similar.

Keyframing is a process in animation that denotes the start and end frames in a time series where parameter values change. To enable analysis of time-lapse images, dNEMO uses a keyframing approach for users to make changes to any user-defined parameter and to track single cells in time-lapse experiments. For example, a user may define a region of the time-lapse where the wavelet map threshold or bounds for acceptable puncta are modified to account for systematic artifacts, and keyframing parameters can be applied in batch across all images in an experiment. A more common use for keyframing in dNEMO is to adjust the segmentation polygon to account for morphology changes and tracking of a cell's movement over the time-lapse image.

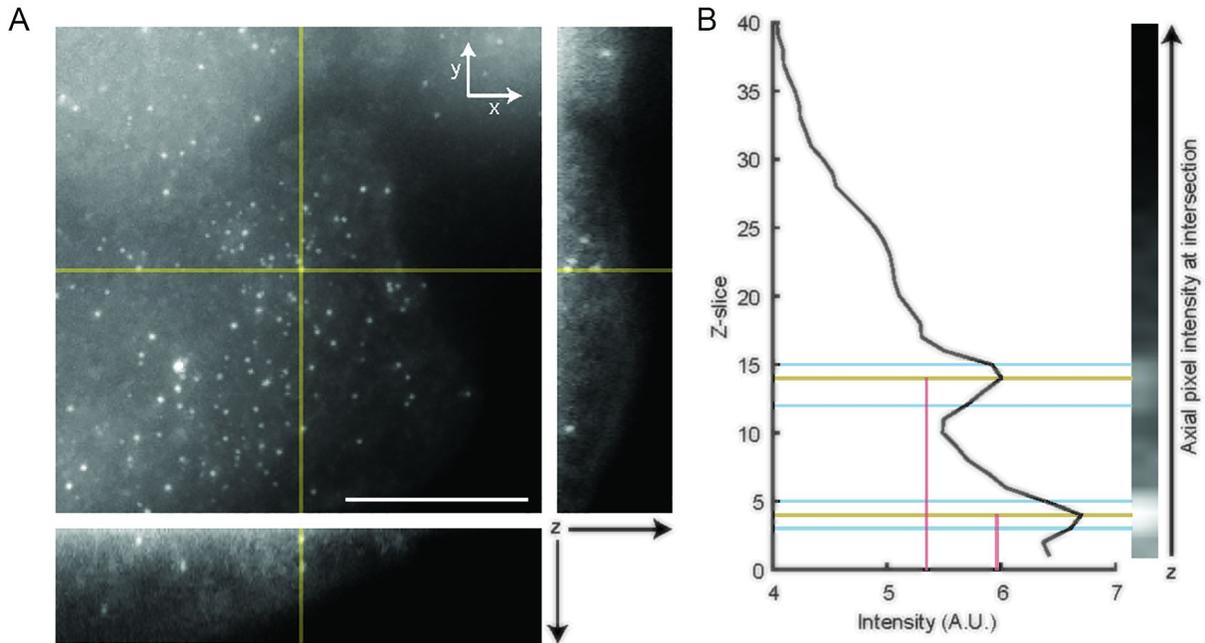
As a demonstration of dNEMO in analysis of 3D time-lapse image, we applied dNEMO to extract kinetic information for SunTag labeled translating polysomes (Figure 2-8). Translating polysomes appearing as diffraction limited GFP puncta in HEK293 cells (Figure 2-3) were identified and counted over time and keyframe adjustment is made to accommodate the mobility of the same cell segmentation polygons. Fluorescent properties of GFP puncta were extracted over the time-lapse to monitor time-courses of corrected intensity, which is a readout of single molecule translation activity [93] and shows only modest photobleaching for extended imaging period. The drop of numbers for translating polysomes suggests a reduce of global translation possibly due to

mRNA degradation. Taken together, distributions for properties of fluorescent puncta can be accurately measured and curated in time-lapse images to produce high-quality datasets in each single cell.



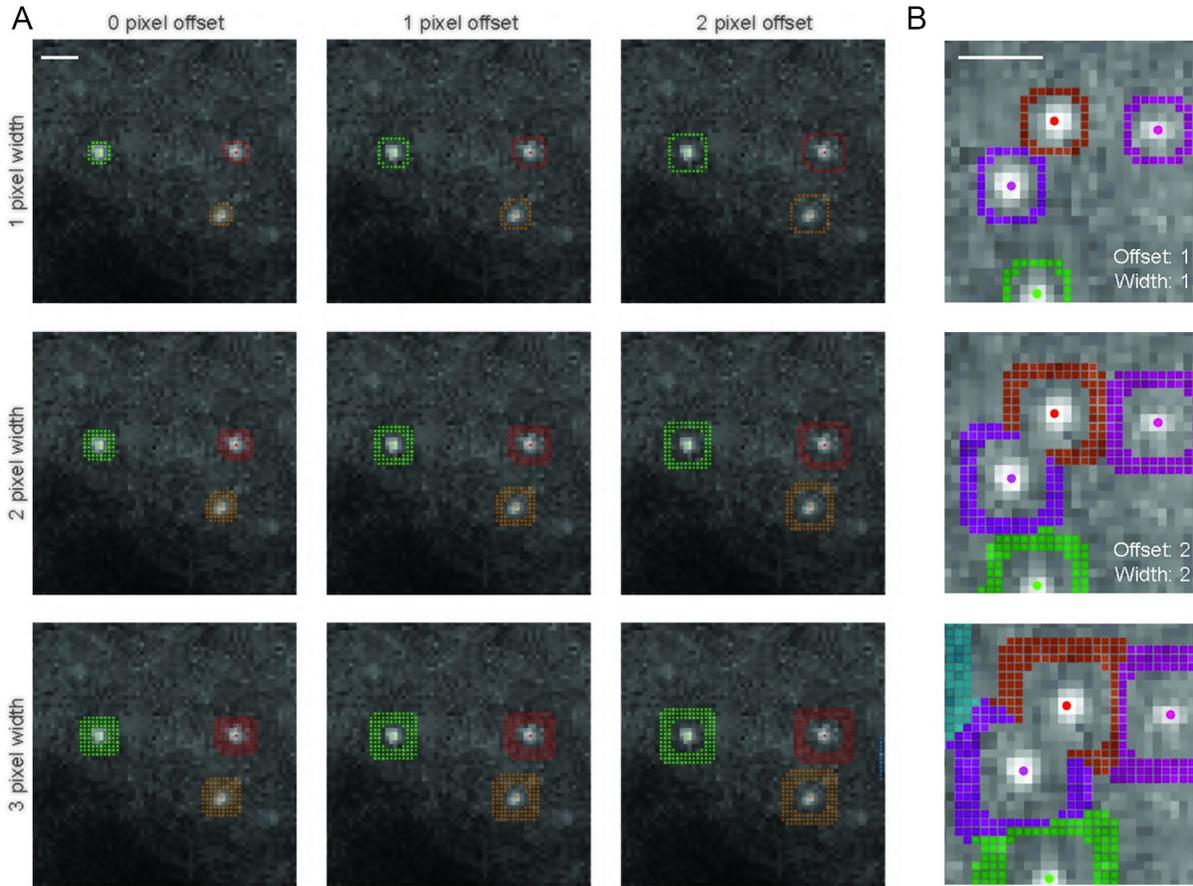
**Figure 2-4** The à trous wavelet transform on simulated and experimental images.

(A) The 3D representations of the convolution matrix (kernel) for levels 1 through 3 of the wavelet transform. (B) Images for simulated data (top), smFISH-labeled NFKBIA transcripts (middle) or GFP-NEMO (bottom), along with the associated L1, L2 and L3 wavelet maps. The L2 wavelet map enhances contrast for diffraction-limited puncta in fluorescence microscopy images. Scale bar: 25  $\mu\text{m}$ . Adapted from [141].



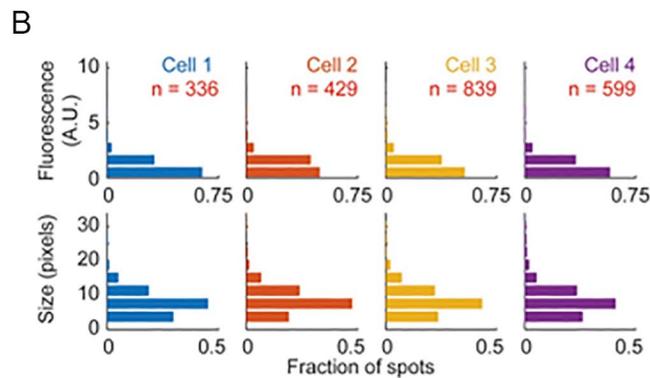
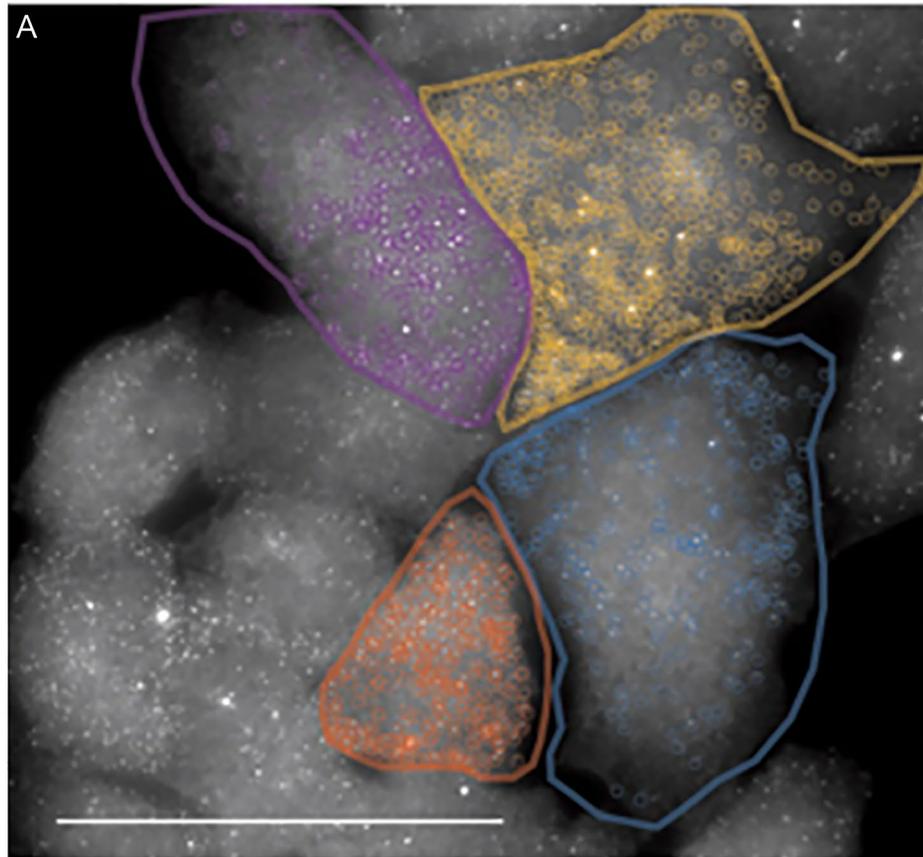
**Figure 2-5 Puncta with axial overlap are individually resolved.**

(A) Maximum intensity projection of a representative NFKBIA smFISH fluorescent z-stack from HeLa cells. Axial projections across the indicated yellow lines for x-z (right) and y-z (bottom) planes show two puncta that overlap in x-y but are axially separated. Scale bar is 10  $\mu\text{m}$ . (B) Axial pixel intensities at the x-y intersection of the yellow lines in panel 'A' (right). The pixel values are measured at each slice of the z-stack (left). Tan lines indicate the axial slice of the centroid for each punctum and teal lines indicate the contiguous z-slices in which the same punctum is observed. Red vertical lines indicate the local background intensity value determined at the axial centroid and used to correct intensity values for each punctum within the 3D image. Adapted from [141].



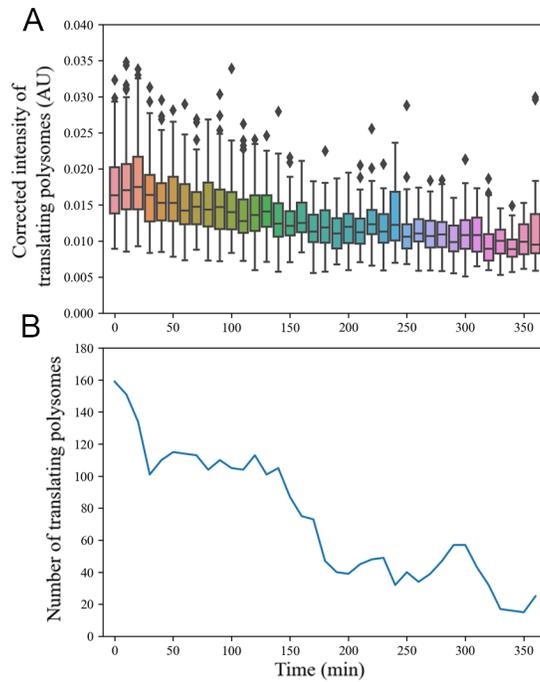
**Figure 2-6 Local background information is used to correct intensity values measured for each punctum.**

(A) For each punctum, pixel intensity values are collected from an annulus with user defined inner radius (offset) and width, defined in pixel units. The mean intensity value of pixels bound by the annulus is subtracted from every pixel within the associated punctum in the source image to correct for the local background. Because each punctum is measured relative to local background, images do not require pre-processing before analysis with dNEMO. (B) Pixels in the vicinity of a spot will be excluded from annuli of all nearby spots. Similarly, when background annuli overlap, background pixels will be associated with the nearest spot only, as determined by its centroid. Images of smFISH against the NFKBIA transcripts in HeLa cells. Scale bars: 1  $\mu\text{m}$ . Adapted from [141].



**Figure 2-7 Identification of smFISH transcripts in fixed-cell images.**

(A) NFKBIA transcripts in HeLa cells labeled by smFISH are identified using dNEMO and associated to single cells. Scale bar 50 microns. (B) Distribution of puncta identified in individual cells reported by intensity (top) or size in number of pixels (bottom). Fluorescence per spot was corrected using the local background about each spot. Adapted from [141].



**Figure 2-8 Quantification of translating polysomes with dNEMO.**

(A) Copy numbers and (B) Boxplots for corrected intensity of translating polysomes identified with dNEMO in a representative cell over time. Corrected intensity is defined as the average of background-corrected pixel values within the area of each detected spot.

## 2.4 Application of dNEMO: Estimating signaling molecule numbers in signaling complex

Signal transmission within the cell is often mediated by formation of punctate structures consisting of a set of signaling molecules [142-145]. As a well-established pathway, NF- $\kappa$ B is the transcription factor that enters nucleus after release from cytoplasm sequestration and activates expression of genes involved in immune and inflammatory responses upon stimulation by inflammatory signals such as tumor necrosis factor (TNF) and interleukin-1 $\beta$  (IL-1). The stimulation starts with binding of TNF to the TNF receptor and recruitment of a group of adaptor proteins and enzymes near the plasma membrane. The enzymes then catalyze the assembly of polyubiquitin scaffolds that recruits the NF- $\kappa$ B essential modulator (NEMO) subunit of the cytoplasmic I $\kappa$ B kinase (IKK) complex and activates IKK to phosphorylate and degrade I $\kappa$ Bs, inhibitor proteins to NF- $\kappa$ B and lead to translocation of NF- $\kappa$ B. IKK complexes appear as punctate structures in this upstream signal transmission process and dNEMO has been used to reveal the size, intensity and dynamical properties of IKK complexes [144] but the exact number of NEMO molecules within each IKK complexes is still unknown.

To quantify the number of NEMO molecules within each IKK complex (spot), the CRISPR labeling of NEMO with EGFP allows us to infer the number of NEMO molecules by counting GFP molecules and converting with 1:1 ratio. In counting GFP molecules in IKK puncta, we used live cell translation reporter developed by Wang et al. [93] and described in Section 2.2 above to calibrate the relation between GFP counts and measured GFP intensity and imaged it in HeLa cells with the same imaging condition used for NEMO in U2OS EGFP-NEMO cells (Figure 2-10).

During translation, one mRNA binds to multiple ribosomes simultaneously and forms a large polysome complex. We assume that the positioning of each ribosome on mRNA independently satisfies a uniform distribution. The signal intensity of each fluorescence foci in the

cytoplasm, representing the translating polysomes, varied depending on the total number of ribosomes and the location of each ribosome on mRNA. When ribosomes reach the region after the coding sequence for the SunTag peptide, the intensity would be the maximum for fully assembled complex; when ribosomes are halfway toward the end, the signal would be approximately half maximal. To account for this variability, we used Monte Carlo simulations to randomly sample the positions of ribosomes and generate the distribution of numbers of SunTag peptides being produced when there are  $n$  ribosomes present on single mRNA, denoted by  $p_2(m, n)$ , where  $m$  is the number of SunTag,  $n$ , as the number of total ribosomes on each mRNA, satisfies Poisson distribution  $p_1(n) = \frac{e^{-\lambda} \lambda^n}{n!}$ . The average number of ribosomes for each translation foci is measured to be 12 experimentally [93]; therefore, we set  $\lambda \approx 12$ . The possibility of  $m$  SunTag peptides being translated on each translation foci are calculated as follows

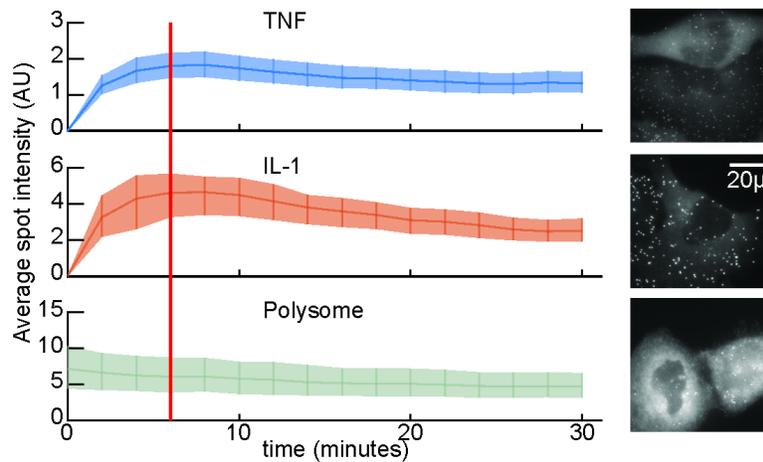
$$P(m) = \sum_{n=1}^{\infty} p_2(m, n) p_1(n)$$

when  $n$  is large, rapidly converges to zero. Therefore, we set  $n = 30$  as the cutoff of the summation.

We quantified the fluorescence intensity of individual translating polysome in HeLa cells using dNEMO with the same threshold used for NEMO quantification in U2OS cells. As the GFP intensity is proportional to the number of GFP molecules, we determined the scaling factor between the measured fluorescence intensity and theoretical distribution of GFP molecule numbers. Here, we used the Nelder-Mead algorithm to minimize the square of difference between the “measured” GFP distribution and theoretical GFP distribution to obtain the scaling factor. To account for effects of photobleaching, we calculated the scaling factors independently for each

frame of the time-lapse image. Using the scaling factor obtained from the translation reporter, the intensity of NEMO spots was converted to numbers of GFP molecules per spot.

The results show that TNF induces complexes that recruit roughly 25% the amount of NEMO-GFP in contrast with IL-1 - although both receptor systems interact with the same cytosolic components. With the quantitative measurement of puncta size and intensity by dNEMO and calibrated GFP numbers with translation reporter, we are able to estimate number of signaling molecules and provide unprecedented resolution to the composition of signaling complexes and facilitate a mathematical understanding of signaling pathways.



**Figure 2-9 Polysome calibration to estimate numbers of EGFP-NEMO molecules per complex.**

U2OS EGFP-NEMO cells stimulated with 100ng/mL IL-1 or TNF were imaged by timelapse microscopy. In parallel, HeLa cells expressing scFv-GFP are transfected with SunTag based translation reporter to reveal polysomes (cells were provided by the Zhuang group and prepared as described in their original work [93]). Imaging conditions were identical for all. Polysome images, with a known number of EGFP molecules per spot were then used to calibrate the intensity of EGFP-NEMO complexes in TNF- or IL-1 stimulated cells at a time when cytokine-induced complexes are brightest (red line). Representative maximum intensity projection images are shown for each. Scale bar: 20 $\mu$ m. Adapted from [144].

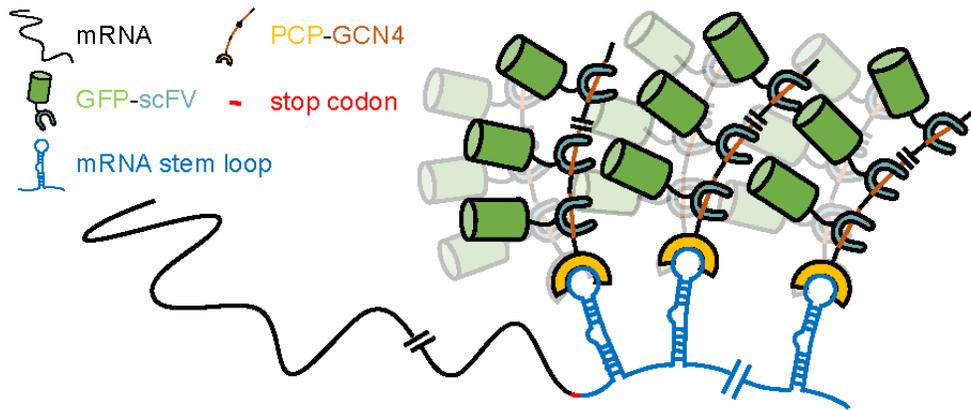
## 2.5 Naïve Design of SunRISER --- Use of SunTag for labeling of single mRNAs

The successful employment of SunTag to observe translation dynamics at single-molecule resolution with standard wide-field microscope demonstrates the power of signal amplification rendered by SunTag. Even the design of SunTag-based translation reporter offers the possibility to simultaneously image single mRNAs in a different fluorescent channel using PP7-PCP system, we and others encounter severe difficulty in resolving single transcripts with standard PP7-PCP using conventional microscopy. We hypothesize that 24 copies of PP7 stem-loops does not provide enough concentration of fluorescent proteins on single mRNAs to make it distinguishable from background and we believe a higher signal amplification enabled by SunTag could be a solution to resolve single transcripts and potentially also facilitate long-term imaging of mRNAs in live cells.

To amplify fluorescence intensity of labeled coat proteins, our design employs SunTag [118], an array of GCN4 peptide epitopes that recruit multiple antibody molecules. The GCN4 antibody-peptide pair bind rapidly and serve as a robust scaffold for protein recruitment with a complex half-life in the order of minutes [146, 147]. Specifically, SunRISER comprises two stages of signal amplification: (1) 24x stem-loop copies are inserted in the 3' UTR of mRNA and (2) the corresponding coat protein is fused with up to 24x SunTag peptides. With co-expression of an FP-fused single-chain antibody (scFv-GFP or codon-optimized scAB-GFP [148]) that bind GCN4 epitopes, each SunTag coat protein can be labeled with up to 24 scFv-FPs, resulting in a theoretical upper bound of over 1,000 FP molecules per mRNA (24x stem loops, 2x coat proteins per stem loop, 24x SunTag arrays,  $24 \times 24 \times 2 = 1,152$  (Figure 2-10)). In contrast with previous approaches, we reasoned that a two-stage approach where SunTag-fused coat proteins and scAB-GFP are both

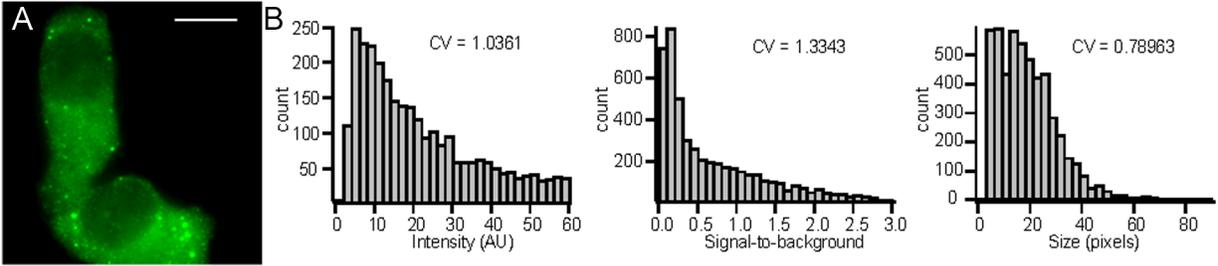
expressed throughout the cell would bolster active cycling of nascent coat proteins and antibody-FPs, providing resistance to photobleaching.

Our initial design for live-cell imaging of mRNAs was first aimed at circumventing problems associated with signal detection and cytoplasmic depletion of FP-fused coat proteins. We expressed components of the naive design in HeLa cells using 24 copies for both PP7 stem loops and SunTag arrays under control of the cytomegalovirus (cmv) promoter and imaged cells by 3D epifluorescence microscopy. As a generic transcript for mRNA detection, detection plasmids expressed a cyan FP (CFP) open reading frame, followed by 24xPP7 inserted after the stop codon in the 3' UTR. Detection plasmids were transfected together with two additional plasmids expressing 24xSunTag-PCP and scFv-GFP. Although diffraction-limited spots were visible, we observed spot-to-spot variability in size and intensity (Figure 2-11).



**Figure 2-10 Design of SunRISER, a SunTag based reporter for imaging signal-enriched mRNA.**

Design schematic of molecular components for the SunRISER strategy to image single mRNA molecules. An mRNA transcript (black) tagged at 3' UTR with PP7 stem-loops (blue) is bound by the PCP coat protein (yellow), which is fused to a SunTag GCN4 peptide array (orange). SunTag recruits GFP (green) through antibody-peptide specific binding between scFv (grey) and GCN4 epitopes. Adapted from [149].



**Figure 2-11 Naïve design shows significant variability in spot intensity and size.**

(A) Maximum intensity projection of z stacks for representative HeLa cells 24 h after transfection with three plasmids: phage-cmv-cfp-24xpp7, cmv-24xSunTag-PCP, and cmv-sfGFP-GB1-scAB and (B) histograms of signal intensity and signal-to-background ratio and size of individual foci quantified with dNEMO. Signal intensity is defined as the average of background-corrected pixel values within the area of each detected spot. Size is measured by number of pixels comprising the spot. Signal-to-background ratio is calculated as the ratio of average pixel intensity within an mRNA spot divided by the average intensity of background pixels in an annular ring surrounding the spot. Inset numbers indicate the coefficient of variation (CV) for each distribution. Adapted from [149].

## 2.6 Concluding remarks

Application of SunTag has enabled imaging translation at the single molecule level in living cells with unprecedented spatial and temporal resolution. This approach allows observation of heterogeneity in the translation properties of different mRNA molecules derived from the same gene in a single cell and preserves spatial translation information that plays an important role in gene regulation in neurons and embryos. We harnessed the remarkable resolution conferred by SunTag and estimated copy numbers of signaling molecules to provide the well-studied NF- $\kappa$ B pathway additional quantitative descriptions. However, naïve implementation of SunTag to label single mRNAs cannot provide single-molecule resolution as expected and shows variability in size and intensity for spot-like structure. A further optimization of SunRISER system is needed to improve detection capability of single mRNAs.

### 3.0 Chapter 3 Development and optimization of SunRISER

This chapter is adapted from:

**Y Guo** and REC Lee, Long-term imaging of individual mRNA molecules in living cells, Cell Reports Methods, 2(6), 100226, 2022, of which I was first author, with the following author contributions:

Conceptualization, Y.G. and R.E.C.L.; methodology, Y.G. and R.E.C.L.; investigation, Y.G.; software, Y.G.; formal analysis, Y.G.; writing – original Draft, Y.G. and R.E.C.L.; writing – review & editing, Y.G. and R.E.C.L.; visualization, Y.G. and R.E.C.L.; funding acquisition, R.E.C.L.; supervision, R.E.C.L.

### 3.1 Modeling of SunRISER system to assist in optimization of SunRISER

The proposed SunRISER design which consists of two-stage signal amplification is expected to enable super bright labeling of mRNAs but naïve attempt without proper modulation yields poor detection of single mRNAs. It is anticipated that multi-stage amplification carries non-linear property and needs meticulous refinement. For example, when scFv-GFP component is excessive and pcp-nxSunTag is limited, each mRNA could be bound by varied numbers of fully assembled pcp-nxSunTag-scFv-GFP complex and shows significant variability in intensity. Another suboptimal scenario is also obvious when superabundant pcp-nxSunTag is present and scFv-GFP is not adequate to occupy all the epitope sites therefore each mRNA will not be labeled a massive amount of GFP as designed. With the complexity brought by two-stage amplification, we constructed a mathematical model to describe SunRISER design quantitatively.

To estimate the number of FP molecules bound to mRNA for different expression levels of SunRISER components, scAB-GFP, SunTag-PCP, and PP7-tagged mRNA, we assume that all dynamic processes, e.g. the expression of proteins, the transcription of mRNAs and the binding/unbinding of proteins, have reached equilibrium. The dissociation constant of scFv and GCN4, can be written as

$$\kappa_{D1} = \frac{[scFv] \cdot [GCN4]}{[scFv \cdot GCN4]} = \frac{(c_1 - p_1 N_1 c_2) \cdot (1 - p_1) N_1 c_2}{p_1 N_1 c_2}$$

where  $c_1$  is the molar concentration scFv.  $c_2$  is the molar concentration of PCP-SunTag.  $N_1$  is the number of GCN4 peptides on the PCP-SunTag.  $p_1$  is the probability of a GCN4 binding site occupied by the scFv. Thus, we obtain

$$p_1 = \frac{1}{2N_1 c_2} \left( \kappa_{D1} + N_1 c_2 + c_1 - \sqrt{(\kappa_{D1} + N_1 c_2 + c_1)^2 - 4c_1 N_1 c_2} \right)$$

Similarly, the probability of a PP7 binding site occupied by a PCP is

$$p_2 = \frac{1}{2N_2c_3} \left( \kappa_{D2} + N_2c_3 + c_2 - \sqrt{(\kappa_{D2} + N_2c_3 + c_2)^2 - 4c_2N_2c_3} \right)$$

where  $c_3$  is the molar concentration of mRNA,  $N_2$  is the number of PP7 stem-loops on the mRNA,  $\kappa_{D2}$  is the dissociation constant between PCP and PP7.

The number of scFv binding to a single PCP-SunTag satisfies a binomial distribution

$$B_1(n) = \binom{N_1}{n} p_1^n (1 - p_1)^{N_1 - n}$$

Similarly, the number of PCP-SunTag binding to a single mRNA molecule satisfies

$$B_2(m) = \binom{N_2}{m} p_2^m (1 - p_2)^{N_2 - m}$$

The average number of scFv-GFP on a single mRNA molecule is thus

$$\bar{n} = \sum_{m=0}^{N_2} \left( m \cdot B_2(m) \cdot \sum_{n=0}^{N_1} n \cdot B_1(n) \right) = N_1 p_1 N_2 p_2$$

To calculate the distribution of number of scFv-GFP binding to a single mRNA molecule, we first calculate the probability of  $n$  scFv-GFP binding to  $m$  PCP-SunTag, which is

$$B_1^m(n) = \binom{mN_1}{n} p_1^n (1 - p_1)^{mN_1 - n}$$

Then the probability of  $n$  scFv-GFP on a single mRNA molecule is the sum of  $B_1^m(n)$  of all possible  $m$ , which is

$$\begin{aligned} P(n) &= \sum_{m=0}^{N_2} B_2(m) \cdot B_1^m(n) = \sum_{m=0}^{N_2} \binom{N_2}{m} p_2^m (1 - p_2)^{N_2 - m} \cdot \binom{mN_1}{n} p_1^n (1 - p_1)^{mN_1 - n} \\ &= (1 - p_2)^{N_2} \left( \frac{p_1}{1 - p_1} \right)^n \sum_{m=0}^{N_2} \binom{N_2}{m} \binom{mN_1}{n} \left( \frac{p_2}{1 - p_2} \right)^m (1 - p_1)^{mN_1} \end{aligned}$$

Although dissociation constants for variant scFv fragments vary significantly, a value of

0.38 nM [150] was selected for scFV-GCN4 in simulations. We note that modeling results showed similar patterns with optimization at a 5:1 ratio of system components across orders of magnitude in parameter sweeps of the dissociation constant for scFV-GCN4. The dissociation constant for PCP/PP7 is 1 nM [151].

Considering that the signal from one mRNA molecule will spread to an ellipsoidal area, of which the size is determined by the Rayleigh radius (Figure 3-1), so we calculate the fluorescence intensity of this area in the presence and without the presence of a single mRNA molecule to calculate signal-to-background ratio. Here, the volume we chose for calculation is  $V=200 \text{ nm} \times 200 \text{ nm} \times 500 \text{ nm}$ . If there is no mRNA in the area, the intensity is defined as the background or noise intensity,  $I_B$ . Assuming that the intensity of single GFP-scFv is 1, the background intensity is the sum of free scFv-GFP molecules and the GFP-SunTag-PCP complex

$$I_B = N_A V (c_1 - N_1 p_1 c_2) + N_A V (c_2 - N_2 p_2 c_3) N_1 p_1$$

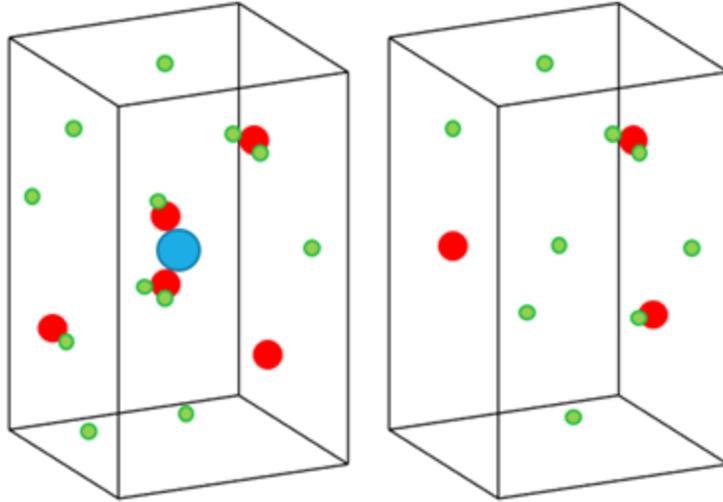
where  $N_A$  is Avogadro constant.

With the intensity of an mRNA molecule as  $I_{RNA} = N_1 p_1 N_2 p_2$  the signal-to-background ratio is

$$\frac{\text{Signal}}{\text{Background}} = \frac{I_{RNA}}{I_B} = \frac{N_1 p_1 N_2 p_2}{N_A V (c_1 - N_1 p_1 c_2) + N_A V (c_2 - N_2 p_2 c_3) N_1 p_1}$$

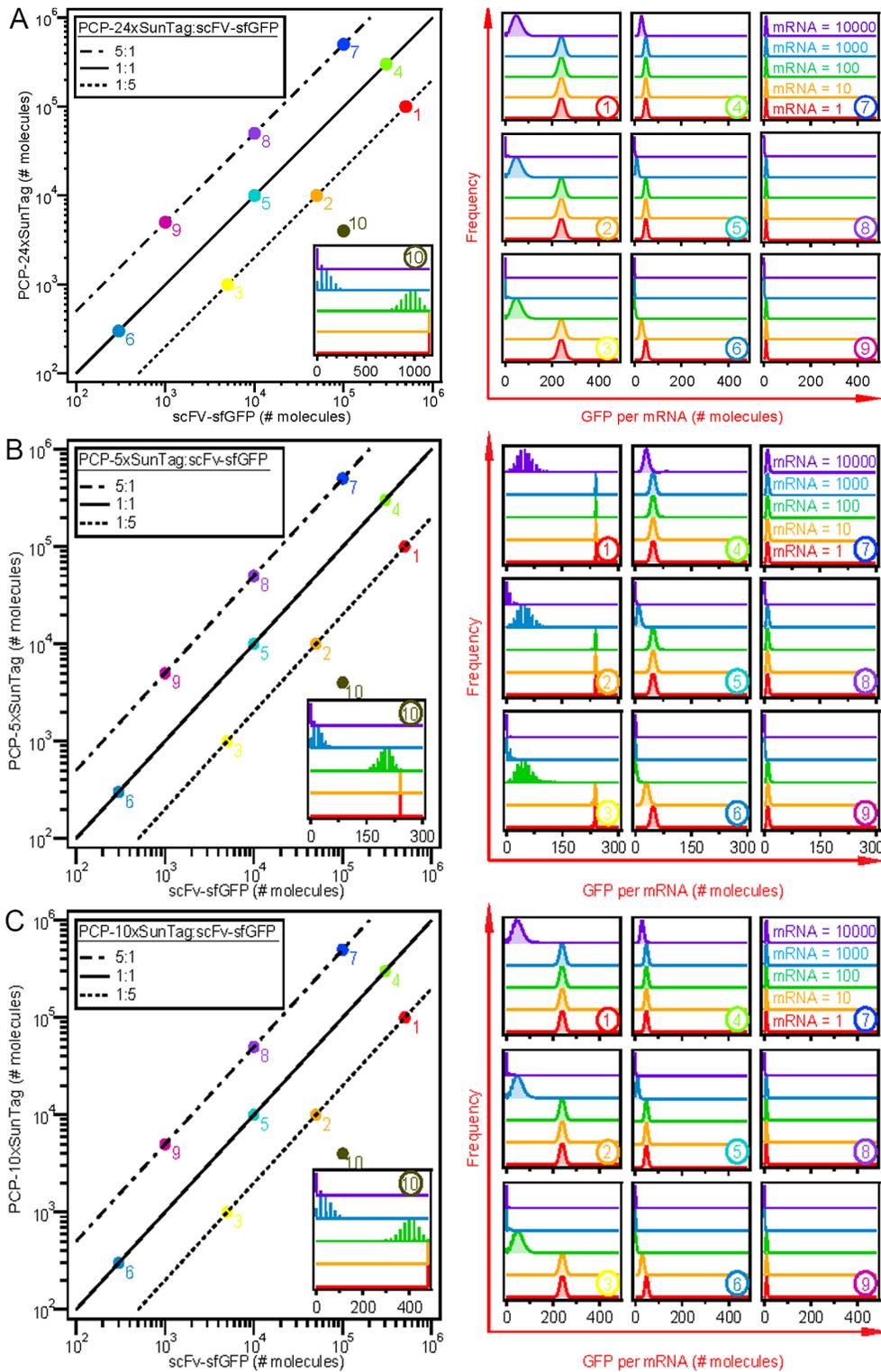
In the sweep of parameter space spanned by number of molecules for scFv-GFP and nxSunTag-PCP, we found broad variability in the expected intensity and signal-to-background (Figure 3-2), in some cases leading to a quantized distribution of single mRNA intensities (e.g., parameter combination 10). Variability between spots will complicate accurate identification and measurement of single mRNA molecules. By inspection, we found that a 5:1 ratio between scAB-GFP and SunTag-PCP with high-abundance expression yields uniform signal intensity distributions in cells expressing up to 1000's of mRNAs (Figure 3-2). We also simulated 5x, 10x

and 24x SunTag-PCP variants. Although longer variants produce more intense signals, all had comparable signal-to-background ratios (Figure 3-3).



**Figure 3-1 The schematic of signal and background definition when using SunRISER to label an mRNA.**

The blue, red and green solid circle represent mRNA, PCP-SunTag and scFv-GFP molecule, respectively. In a volume which is defined by Rayleigh radius of the microscope, the background is defined as the sum of the number of free scFv-GFP and the number of scFv-GFP on the free PCP-SunTag. The signal is defined as the mean number of scFv-GFP on a single mRNA molecule.

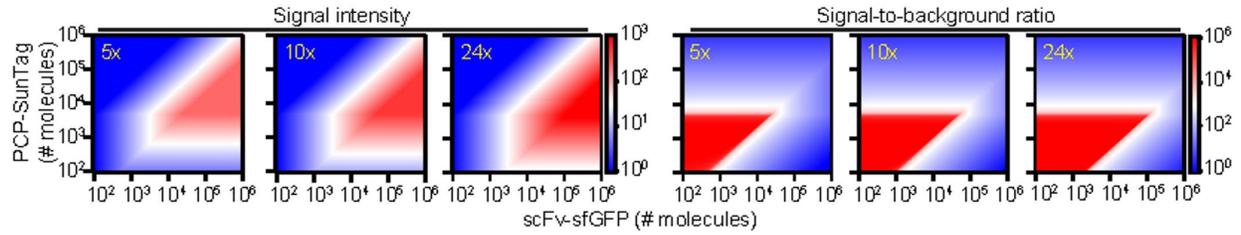


**Figure 3-2** Parameter sweeps in the space spanned by number of molecules for scFv-GFP and nxSunTag-PCP

(left, (A)  $n=24$  (B)  $n=5$  (C)  $n=10$ ) using a computational model to calculate number of GFP molecules per

mRNA.

Parameter combinations 1-10 were selected to represent different ratios between scFv-GFP and nxSunTag-PCP (solid line 1:1, dotted line 5:1 and dash-dotted line 1:5) and different expression levels and frequency plots are shown (right). Different concentrations of mRNAs per cell are distinguished by different colors. Adapted from [149].



**Figure 3-3 Heatmaps of mean number of GFP molecules per mRNA (left) and signal-to-background ratio (right) in the parameter space spanned by the number of molecules for scFv-GFP and ‘n’xSunTag-PCP (from left to right, 5x, 10x and 24x).**

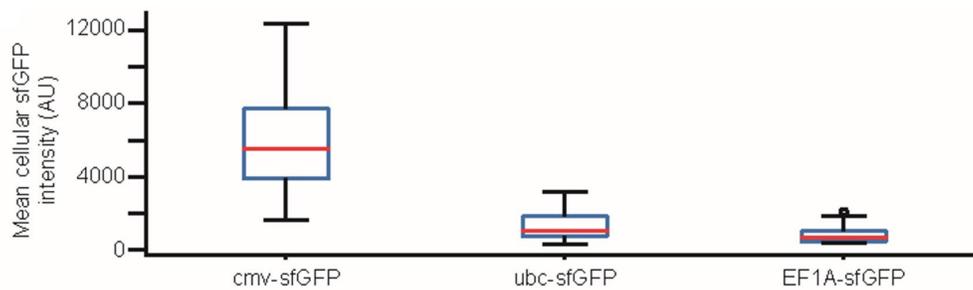
Signal-to-background ratio is calculated as ratio of scFV-sfGFP molecules bound to an mRNA divided by the average intensity of unbound scFV-sfGFP in the background. Heatmaps were calculated assuming 100 mRNA molecules per cell. Adapted from [149].

### 3.2 Experimental validation of optimal expression ratio predicted by model

Guided by simulations, we designed SunRISER variants and assayed quantitative properties of mRNA spots. The expression level we have described in the mathematical model refers to numbers of protein molecules, which are the final product of gene expression process. Each step in the transfer of information from DNA to RNA to protein can potentially be regulated to adjust the amount of proteins one cell produces, which is determined by the balance between biochemical processes of protein synthesis and degradation. Transcription as the initiation step of protein production is the primary efficient control for protein levels as one mRNA can be translated to many copies of protein. Transcription machinery involves different types of RNA polymerase (RNAP) and they recognize specific regulatory sequences upstream of target genes to decide when and where a specific gene would be expressed. One major type of regulatory sequences is called promoters, the strength of which binding to RNAP and transcription factors dictates the rate of transcription and therefore mRNA copies present in the cell. Experimental studies have established libraries of promoters with different strengths that are often employed in synthetic biology to achieve required level of protein expression.

To establish an approximately 5:1 ratio of abundance for SunRISER protein components, we compared different constitutive promoters in HeLa cells (Figure 3-4) predicted to have different strengths [152]. Comparing CMV and *ubc*, promoters for strong and weak (respectively) mammalian gene expression, we found that the abundance of GFP expressed from the *cmv* promoter is approximately 5 times that of *ubc* (Figure 3-4).

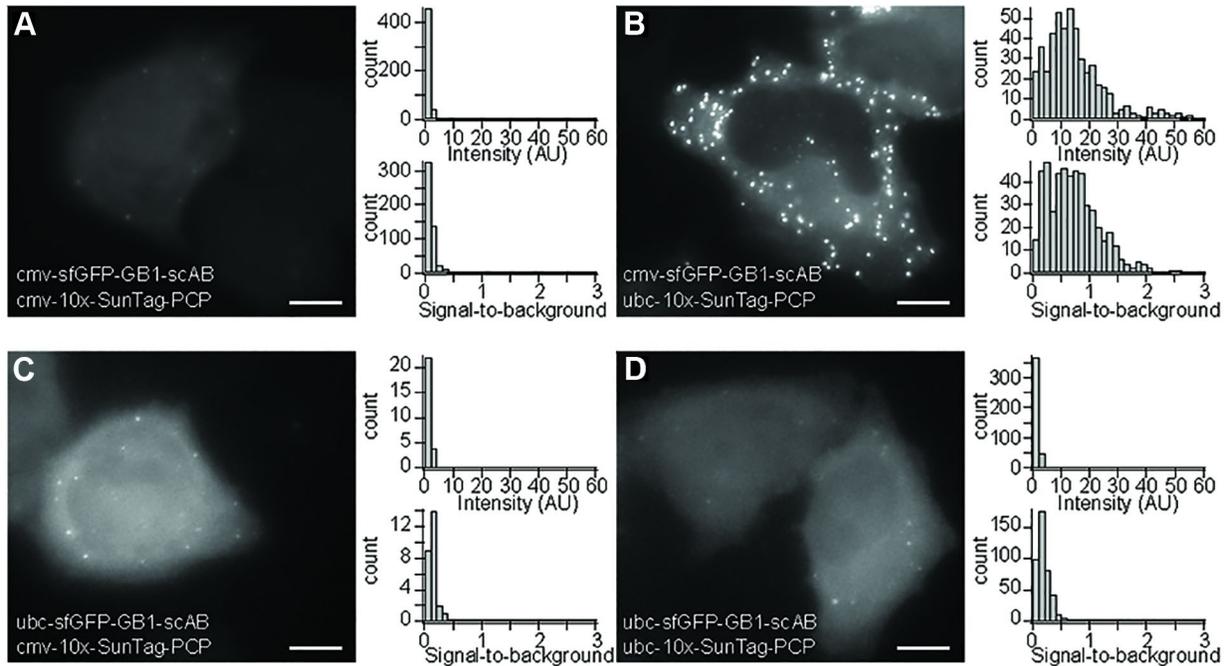
With promoters selected to approximate the desired expression ratio predicted from simulation, we implement SunRISER design with *cmv* driving expression of scFv-GFP and *ubc* driving expression of pcp-nxSunTag. To reduce the size of labeled mRNA complexes without



**Figure 3-4 Quantification of strengths for constitutive promoters in HeLa cells.**

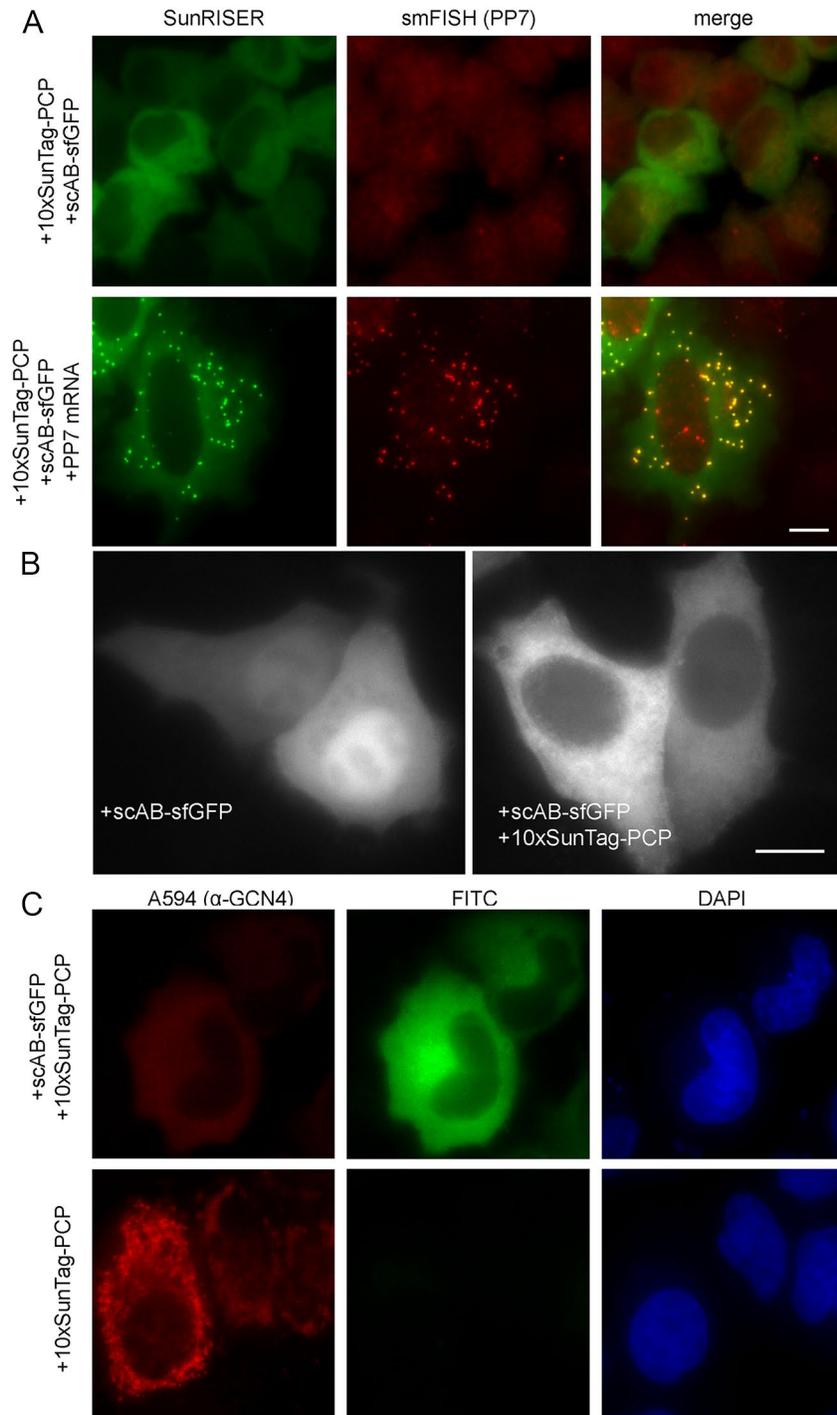
HeLa cells transfected with cmv-sfGFP-GB1-scAB, ubc-sfGFP-GB1-scAB, or EF1A -sfGFP-GB1-scAB were imaged 24 hours after transfection and mean GFP fluorescence intensity was measured in single cells. Boxplots (median and interquartile ranges) show expression variability for indicated promoters. The expression of GFP driven by the cmv promoter is approximately 5-times and 7-times that by ubc and EF1A promoters respectively. Adapted from [149].

significantly compromising signal to background, we validated model predictions using 10xSunTag-PCP (Figure 3-3). HeLa cells were co-transfected with SunRISER components, approximating parameter combinations 1, 4, 6, and 9 (Figures 3-2). Consistent with simulations, an approximately 5:1 ratio achieved by cmv-scAB-GFP and ubc-SunTag-PCP (Figure 3-5B) enhanced signal intensity and signal-to-background values compared to suboptimal ratios (Figure 3-5A, C, D) that fail to reliably and unambiguously label single transcripts driven by the same CMV promoter. Using smFISH against PP7 stem-loop sequences [153] in SunRISER-expressing cells, we confirmed co-localization of cytoplasmic mRNAs between SunRISER and smFISH. However, smFISH revealed nuclear mRNAs that were not detected by SunRISER (Figure 3-6A). Further live-cell and fixed-cell assays demonstrated that GFP-SunTag-PCP is excluded from the nuclear compartment (Figure 3-6B, C), suggesting that SunTag-PCP required further optimization for whole-cell mRNA detection.



**Figure 3-5 Varying promoters for scAB-GFP and 10xSunTag-PCP corroborates model predictions.**

HeLa cells transfected with phage-cmv-cfp-24xpp7 and indicated constructs were imaged by 60x wide-field microscopy 24 hours after transfection and imaged with identical settings, representative maximum intensity projections are shown. Images were analyzed with dNEMO to quantify mRNA spots. Signal intensity is defined as the average of background-corrected pixel values within the area of each detected spot. Signal-to-background ratio is calculated as ratio of signal intensity divided by the average intensity of background pixels in an annular ring surrounding the spot. Note that images in (C) and (D) were contrast-enhanced for visualization. The combination of cmv-sfGFP-GB1-scAB and ubc-10xSunTag-PCP enabled robust detection of diffraction-limited mRNA spots and shows the highest signal intensity and signal-to-background ratio. Scale bar: 10  $\mu$ m. Adapted from [149].



**Figure 3-6**The preliminary version of SunRISER labels only cytoplasmic mRNA because SunTag-PCP is excluded from nuclei.

(A) Maximum intensity projection images of smFISH performed with probes against the pp7 stem-loops on HeLa cells transfected with indicated constructs are shown. The combination of cmv-

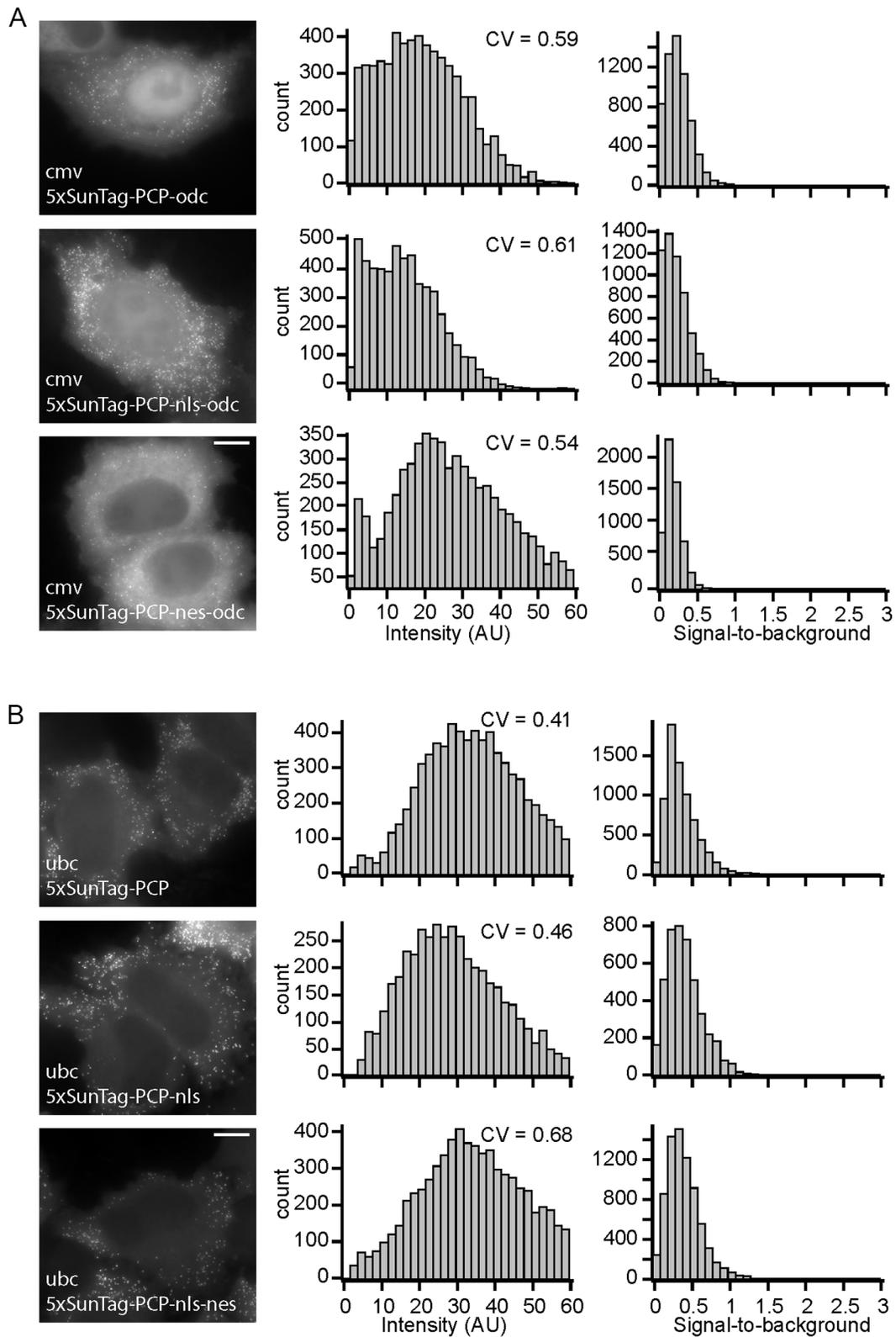
sfGFP-GB1-scAB and ubc-10xSunTag-PCP allows visualization of cytoplasmic mRNA, but not mRNAs in nucleus. From left to right, FITC CY5, and merged channels are shown. (B) HeLa cells were transfected with cmv-sf-GFP-GB1-scAB only (left) or co-transfected with a combination of cmv-sfGFP-GB1-scAB and ubc-10xSun-Tag-PCP (right) and imaged after 24 hours. Co-expression of scAB-GFP and SunTag-PCP depletes scAB-GFP from the nucleus through interaction with SunTag-PCP. (C) HeLa cells transfected with cmv-sfGFP-GB1-scAB and ubc-10xSunTag-PCP (top) and ubc-10xSunTag-PCP only (bottom) were stained for SunTag using a  $\alpha$ -GCN4 antibody. Since sfGFP-scAB competes with the  $\alpha$ -GCN4 antibody for the GCN4 epitope, co-expressing cells are lower intensity in the A594 channel (top left). In both conditions, 10xSunTag-PCP is predominantly cytoplasmic regardless of whether sfGFP-scAB is expressed in the same cell. Scale bars: 10  $\mu$ m for all. Adapted from [149].

### 3.3 Refinement of SunRISER for whole-cell mRNA imaging

To alleviate nuclear export effects from repeats of the GCN4 epitope and cytoplasmic sequestration of anti-GCN4 [150], we continued with the smaller 5xSunTag variant. Next, we focused on modifications to 5xSunTag-PCP for homogeneous expression throughout the cell and detection of nuclear and cytoplasmic mRNAs. We also considered that ornithine decarboxylase (ODC) tag [154] fused to SunTag-PCP while under control of the *cmv* promoter as an alternative approach to establish a 5:1 deficit of SunTag-PCP that may also alter its sub-cellular distribution. ODC is one of the most short-lived proteins, which can facilitate fast degradation of tagged proteins. Although ODC tagging is not as precise as controlling transcript abundance using promoters, ODC fusion of SunTag-PCP is a viable approach to limit its expression relative to scAB-GFP. Although several variants enable whole-cell mRNA detection, an optimized design was eventually achieved by switching the fusion order of PCP and 5xSunTag, in addition to inserting a 5' NLS (Figures 3-7 and Table 3-1). We note that addition of a 5' NLS without switching the fusion order to PCP-SunTag still fails to detect nuclear mRNAs when expressed with other SunRISER components.

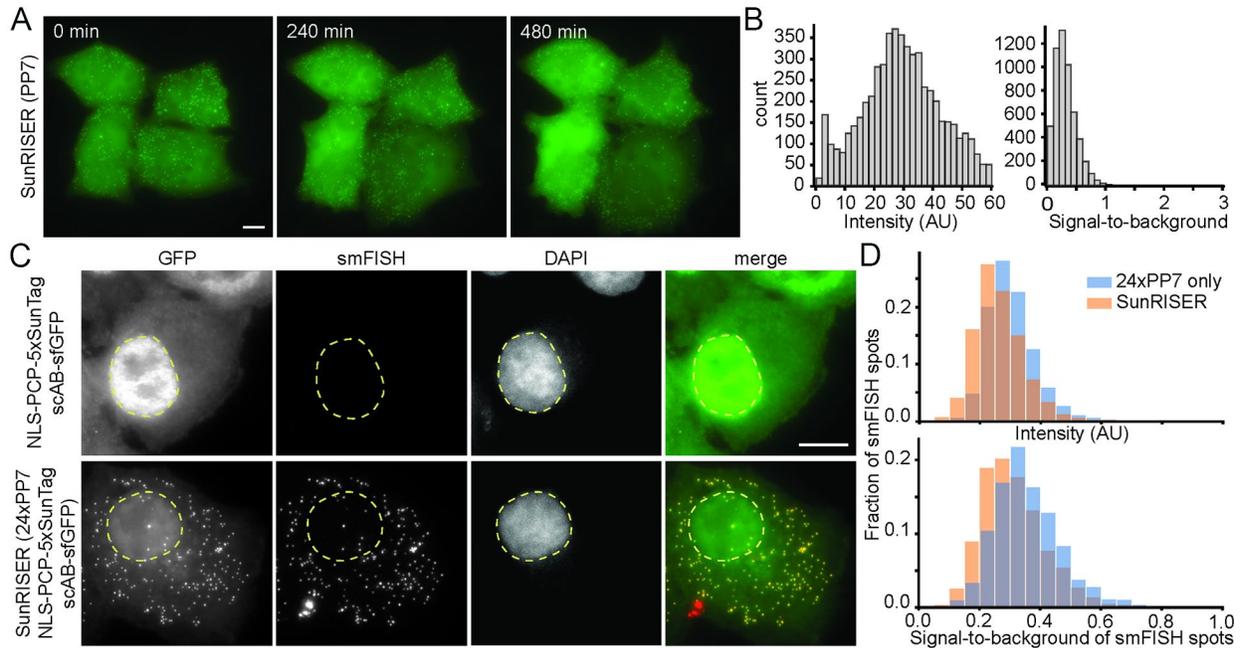
The optimized SunRISER version 1 (SRv1) design consists of three plasmids (i: *cmv*-GFP-GB1-scAB, ii: phage-nls-PCP-5xSunTag, and iii: a detection plasmid expressing mRNA tagged with 24XPP7) and has an overall molecular weight comparable with MS2x128 (Table 3-2). SunRISER faithfully labels single mRNAs in the nucleus and cytoplasm with uniform fluorescence intensity and high signal-to-background, allowing for long term imaging (Figures 3-8). We also compared fluorescence distributions of single cytoplasmic mRNAs detected by smFISH in cells expressing either the detection plasmid only or the complete SunRISER system. We found that expression of SunRISER components does not significantly alter the fluorescence

intensity and signal-to-background of mRNA spots (Figure 3-8D), suggesting that each SunRISER spot represents a single mRNA molecule. Similarly, detection plasmids using weak and strong promoters to express low and high mRNA numbers result in expected mRNA abundances, consistent fluorescence intensity, and signal-to-background, in-line with expectations from the model (Figure 3-9A, B, C). For some cell lines, calibration of promoters for protein components may be necessary to ensure optimal SunRISER labeling. Nevertheless, even though promoter activity will vary by cell type, expression of SRv1 components in A549 cells produced qualitatively similar results even though A549 cells are phenotypically distinct from HeLa by many criteria (Figure 3-9D). Furthermore, the same optimization can be applied to orthogonal stem-loops and antibody-epitope pairs, such as MS2, MS2V6 [107], and MoonTag [119] (Figures 3-10 and 3-11).



**Figure 3-7 Modifications to SunTag-PCP alter sub-cellular distribution of SunRISER components.**

(A)-(B) HeLa cells transfected with cmv-sfGFP-GB1-scAB, phage-cmv-cfp-24xpp7 and indicated constructs were imaged with identical settings by 60x wide-field microscopy 24 hours after transfection and quantified with dNEMO. Maximum intensity projections of representative cells are shown. Signal intensity is defined as the average of background-corrected pixel values within the area of each detected spot. Signal-to-background ratio is calculated as ratio of signal intensity divided by the average intensity of background pixels in an annular ring surrounding the spot. Coefficient of variation (CV) is the ratio of the standard deviation to the mean. ODC acts as a degron to reduce the expression level of SunTag-PCP driven by cmv promoter; nls, nuclear localization signal; nes, nuclear export signal. Scale bar: 10  $\mu$ m. See also Table 3-1. Adapted from [149].

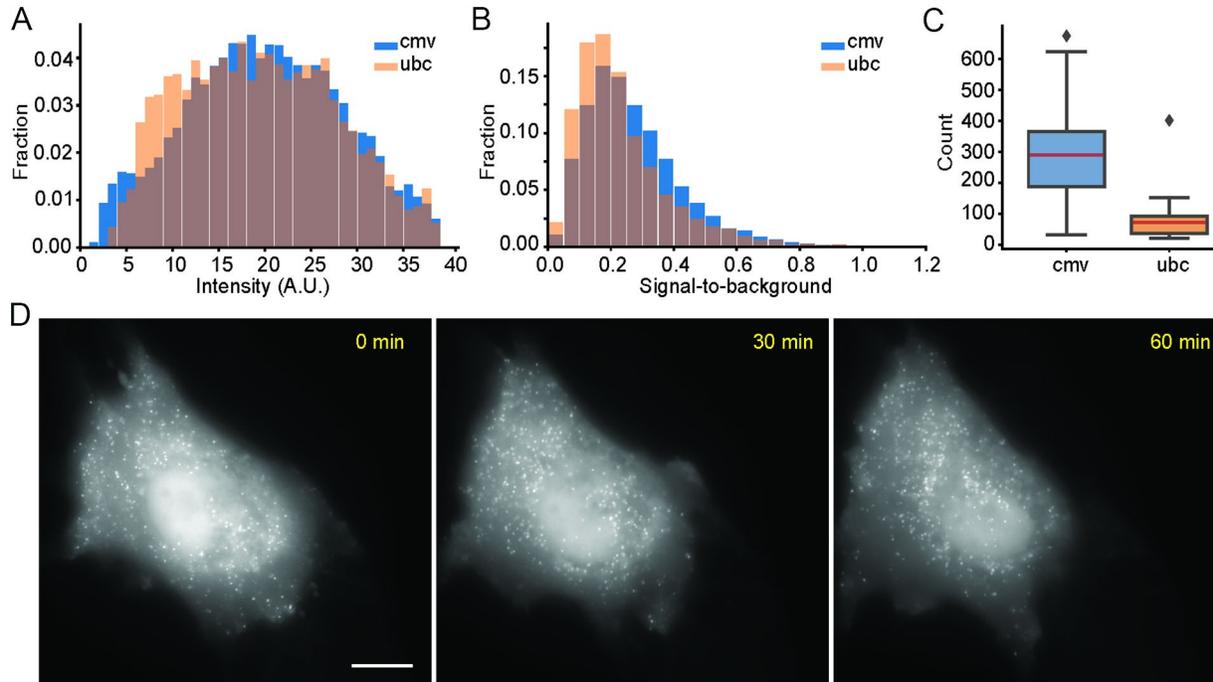


**Figure 3-8 The optimal SunRISER design allows long-term imaging of single mRNA molecules throughout the cell.**

(A) Maximum intensity projections of HeLa cells transfected with SunRISER, consisting of the detection plasmid phage-cmv-cfp-24xpp7, as well as *ubc-nls-PCP-5xSunTag* and *cmv-sfGFP-GB1-scAB*. Cells were imaged by 60x widefield fluorescence microscopy for 8 hours with a 10-minute framerate. Scale bar: 10  $\mu\text{m}$ . (B) Histograms of signal intensity and signal-to-background ratio of SunRISER labeled single mRNAs quantified with dNEMO. Signal intensity is defined as the average of background-corrected pixel values within the area of each detected spot. Signal-to-background ratio is calculated as the ratio of average pixel intensity within an mRNA spot divided by the average intensity of background pixels in an annular ring surrounding the spot. ( $n=23$  for cell numbers and  $n=5611$  for spots numbers). (C) Maximum intensity projection images of smFISH performed with probes against the PP7 stem-loops in HeLa cells transfected with SunRISER protein components (*ubc-nls-PCP-5xSunTag* and *cmv-sfGFP-GB1-scAB*) only (top) or complete SunRISER with detection plasmid (bottom). Scale bar: 10  $\mu\text{m}$ . (D) Histograms of

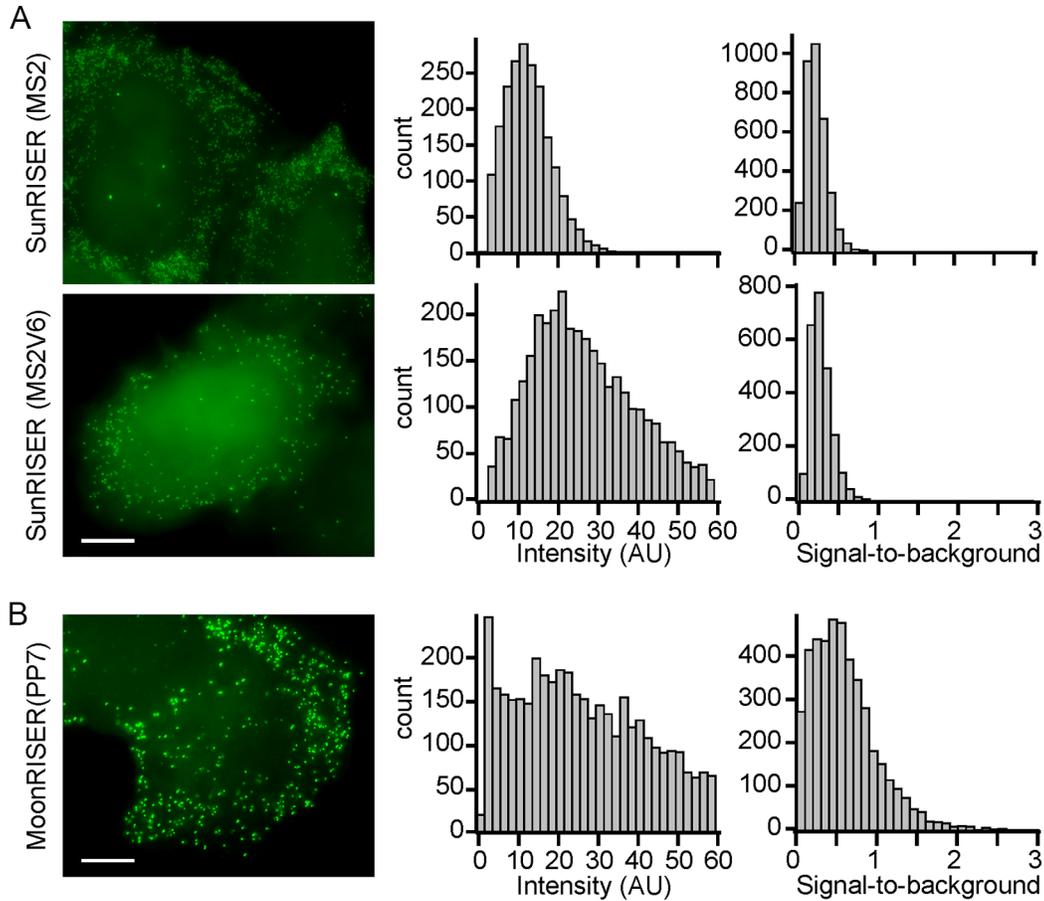
signal intensity and signal-to-background ratio for single mRNAs labelled with smFISH Q670 probes against the mCherry coding sequence in cells expressing phage-cmv-mCherry-24xpp7 detection plasmid without (blue, n=10 for cell numbers and n=2156 for spots numbers) and with (orange, n=11 for cell numbers and n=2842 for spots numbers) remaining SunRISER components.

Adapted from [1].



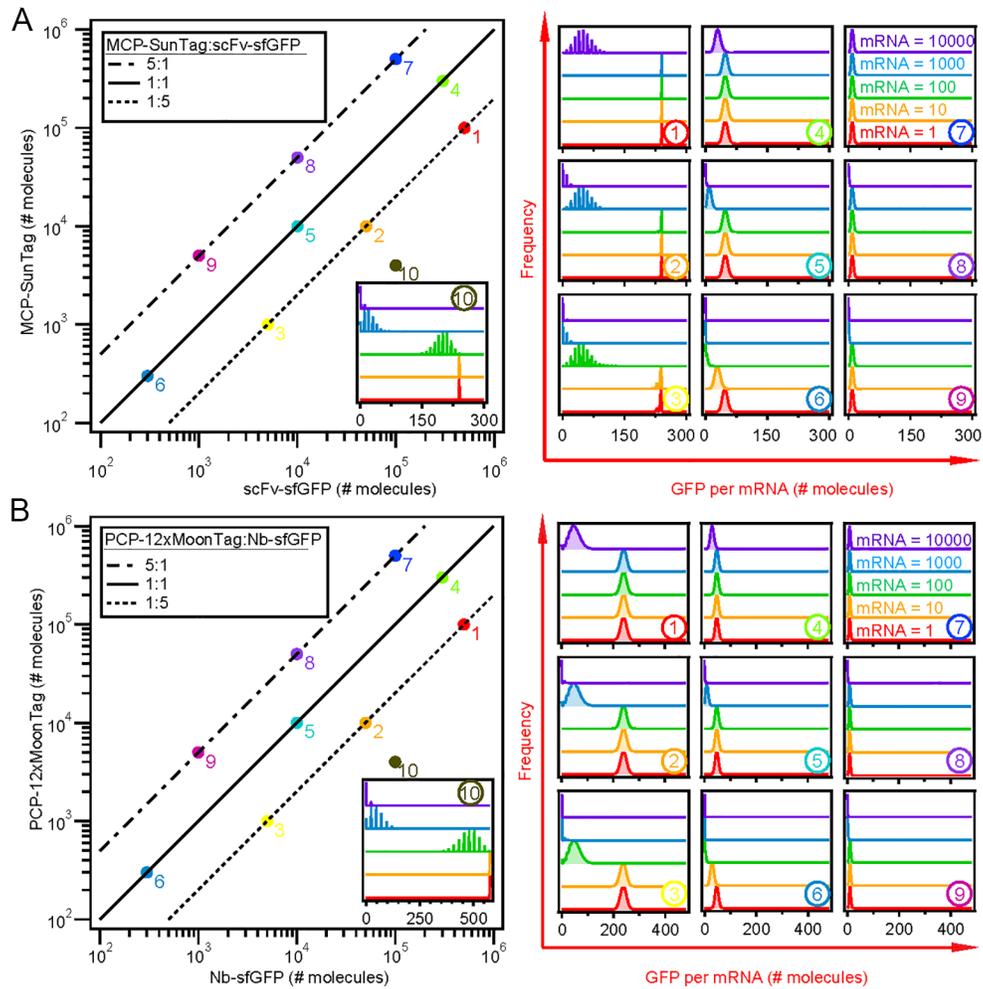
**Figure 3-9 SunRISER-labeled mRNAs comparing weak and strong promoters in the detection plasmid, and expression in A549 cells.**

(A-C) Comparison of signal intensity (A), signal-to-background ratio (B), and mRNA numbers (C), between mRNAs expression driven by *cmv* and *ubc* promoters in the detection plasmid. Blue bars are *cmv*-driven and orange bars are *ubc*-driven SRv1-labeled mRNA. (D) Maximum intensity projection images of A549 cells transfected with SunRISER SRv1 phage-*cmv*-*cfp*-24xpp7, *ubc*-nls-PCP-5xSunTag and *cmv*-sfGFP-GB1-scAB are shown. Note that A549 cells were imaged with longer exposure times to HeLa for similar overall image intensity, possibly related to the large size of A549 cells. Scale bar: 10  $\mu$ m. Adapted from [149].



**Figure 3-10 Design principles of SunRISER are generalizable to orthogonal stem-loops and protein tagging systems.**

(A) HeLa cells transfected with detection plasmids phage-cmv-CFP-24xMS2 (top) and phage-cmv-CFP-24xMS2V6 (bottom) stem loops variants of SunRISER with *ubc-nls-MCP-5xSunTag* show similar characteristics and intensity distributions, quantified in histograms (right). (B) HeLa cells transfected with detection plasmid phage-cmv-cfp-24xpp7 with *cmv-sfGFP-GB1-Nb-gp41*, and *ubc-nls-PCP-12xMoonTag*, quantified in histograms (right). We note that the MoonRISER example can be further optimized as it uses a longer 12x MoonTag and a nanobody which has different binding properties. Cells were imaged with 60x wide-field microscope 24 hours after transfection and quantified with dNEMO. Scale bar: 10  $\mu\text{m}$ . Adapted from [149].



**Figure 3-11 Model calibrated to orthogonal RNA hairpin and coat protein systems are comparable with previous observations for PP7.**

(A) Parameter sweeps in the space spanned by the number of molecules for scFV-GFP and 5xSunTag-MCP (left) using a computational model to calculate number of GFP molecules per mRNA. Parameter combinations 1-10 were selected to represent different ratios between scFV-GFP and 5xSunTag-MCP (solid line 1:1, dotted line 5:1 and dash-dotted line 1:5) and different expression levels and frequency plots are shown (right). Different concentrations of mRNAs per cell are distinguished by different colors. (B) Parameter sweeps in the space spanned by the number of molecules for Nb-gp41-GFP and 12xMoonTag-PCP (left) using a computational model to calculate number of GFP molecules per mRNA. Parameter combinations 1-10 were selected to

represent different ratios between Nb-gp41-GFP and 12xMoonTag-PCP (solid line 1:1, dotted line 5:1 and dash-dotted line 1:5) in plots of frequency versus intensity of GFP labeling (right). Different concentrations of mRNAs per cell are distinguished by different colors. Adapted from [149].

**Table 3-1 Summary of SunTag/PCP plasmid variants.**

PCP/SunTag variants	Nuclear GFP	Nuclear spots	Cytoplasmic spots
<b>ubc-10xSunTag-PCP-nls</b>	No	Limited	Yes
<b>ubc-10xSunTag-PCP-nls-nes</b>	No	Limited	Yes
<b>cmv-10xSunTag-PCP-ODC</b>	Yes	No	Yes
<b>cmv-10xSunTag-PCP-nls-ODC</b>	Yes	Yes	Yes
<b>cmv-10xSunTag-PCP-nes-ODC</b>	Yes	Limited	Yes
<b>cmv-10xSunTag-PCP-nls-nes-ODC</b>	Yes	Limited	Yes
<b>ubc-10xSunTag-PCP-ODC</b>	Limited	No	Limited
<b>ubc-10xSunTag-PCP-nls-ODC</b>	Yes	Limited	Limited
<b>ubc-10xSunTag-PCP-nes-ODC</b>	Limited	No	Limited
<b>ubc-10xSunTag-PCP-nls-nes-ODC</b>	Yes	No	Limited
<b>cmv-5xSunTag-PCP-ODC</b>	Limited	No	Yes
<b>cmv-5xSunTag-PCP-nls-ODC</b>	Yes	No	Yes
<b>cmv-5xSunTag-PCP-nes-ODC</b>	No	No	Yes
<b>ubc-5xSunTag-PCP-nls</b>	No	Limited	Yes
<b>ubc-nls-5xSunTag-PCP</b>	Limited	Limited	Yes
<b>ubc-5xSunTag-PCP-nls-nes</b>	No	Limited	Yes
<b>ubc-nls-5xSunTag-PCP-nls</b>	Yes	No	Yes
<b>ubc-nls-5xSunTag-PCP-2xnls</b>	Yes	No	Limited
<b>ubc-2xnls-5xSunTag-PCP-2xnls</b>	Yes	No	No
<b>ubc-PCP-5xSunTag</b>	Limited	Limited	Yes
<b>ubc-nls-PCP-5xSunTag</b>	Yes	Yes	Yes
<b>ubc-nls-PCP-5xSunTag-nes</b>	Limited	Limited	Yes

SunTag-PCP sequence modifications include varying promoters (ubc and cmv), nuclear localization sequences (nls), nuclear export sequences (nes) and ornithine decarboxylase (ODC).

Note that “Limited” indicates that significant cell-to-cell variability or lower number than expected was observed for the associated property. Adapted from [149].

**Table 3-2 Molecular weight for SunRISER variants and other mRNA labeling techniques for labeling single mRNAs.**

<b>Variants</b>	<b>Molecular weight of mRNA stem-loops</b>	<b>Molecular weight of protein cargos</b>	<b>Total molecular weight</b>
<b>24xPP7</b>	426.34kDa	2100.48kDa	2526.82kDa
<b>MS2x128</b>	2233.6kDa	10931.2kDa	13164.8kDa
<b>SunRISER-V1 (24xPP7-5xST)</b>	426.34kDa	15528.48kDa	15954.82kDa
<b>SunRISER-V1.1 (8xPP7-10xST)</b>	134.86kDa	10085.76kDa	10220.62kDa
<b>SunRISER-V1.2 (10xPP7-12xST)</b>	154.17kDa	15062kDa	15216.17kDa

Molecular weight calculations for fully assembled mRNA labeled complexes, including associated coat proteins fused to GFP. Calculated weight for mRNA includes only the stem loops and does not consider an associated ORF. RNA molecular weight is calculated with Quest Calculate™ RNA Molecular Weight Calculator [155] and protein molecular weight is calculated with The Sequence Manipulation Suite [156].

**Table 3-3 Summary of parameters used in simulation.**

	Kd of antibody-peptide pair (mol/L)	Kd of coat protein and stem-loop pair (mol/L)	# of epitopes on coat protein
<b>PCP-5xSunTag +mRNA-24xPP7</b>	0.38e-9	1e-9	5
<b>PCP-10xSunTag +mRNA-24xPP7</b>	0.38e-9	1e-9	10
<b>PCP-24xSunTag +mRNA-24xPP7</b>	0.38e-9	1e-9	24
<b>MCP-5xSunTag +mRNA-24xMS2</b>	0.38e-9	1e-9	5
<b>MCP-5xSunTag +mRNA-24xMS2V6</b>	0.38e-9	2.4e-9	5
<b>PCP-12xMoonTag +mRNA-24xPP7</b>	30e-9	1e-9	12
<b>Common constants</b>	Volume of whole cell = 1745 $\mu\text{m}^3$	Volume of Rayleigh criterion = 0.02 $\mu\text{m}^3$	# of stem-loops = 24

List of tunable parameters used in models simulating different variants of SunRISER with orthogonal stem-loops and antibody-epitope pairs. The volume of HeLa cell is from Bionumbers database (ID 103725), which can be adjusted for different cell types or modified for use with other measurements of cell volume. The volume of Rayleigh criterion is calculated as 200nmx200nmx500nm (lateral resolution for light microscope is 200nm and axial resolution is 500nm). Note that Kd for anti-GCN4 scFv can range from  $10^{-6}$  to  $10^{-11}$  M.

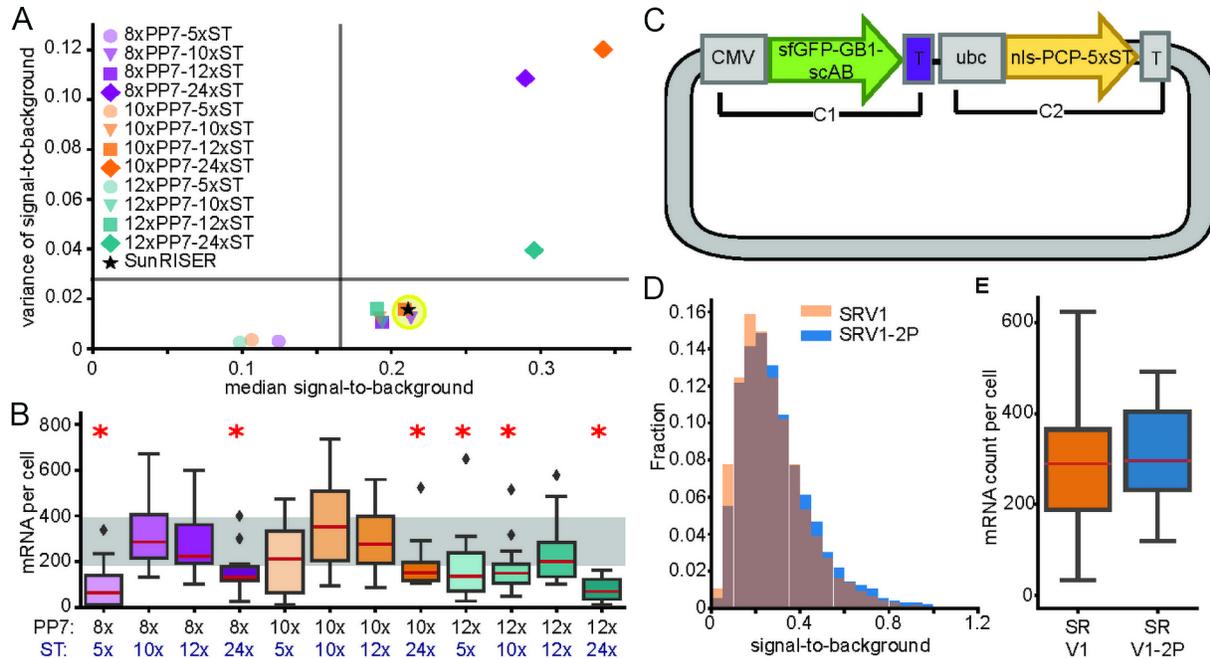
### 3.4 SunRISER variants with small stem-loop arrays and reduced plasmid requirements

Most mRNA-tagging applications using bacteriophage-derived stem-loops use a 24x copy array or larger for signal amplification. Shorter stem-loop arrays exist, but are often used in CRISPR based genome imaging, where multiple copies of target sites are present to detect bright foci, or engineered to reduce impact on general mRNA metabolism [157-159]. We therefore asked whether SunRISER can be used to label single mRNAs using shorter stem-loop arrays.

To establish versatility of our approach, we developed SunRISER configurations using shorter PP7 stem-loop arrays (8x, 10x, 12x), and varying lengths of PCP-SunTag arrays (5x, 10x, 12x, and 24x). For each array combination, we measured median and variance of signal-to-background for distributions of single mRNA molecules (Figure 3-12 and 3-13). We found that although longer SunTag arrays tend to correspond with greater signal-to-background, they also result in greater inter-spot variance which reduces mRNA detection efficiency (Figure 3-12A and B). Of all combinations tested, we found two SunRISER variants capable of labeling mRNAs with low variance, comparable signal-to-background, and comparable mRNA detection numbers to SunRISER (Figures 3-12 and 3-13). We refer to these as SRv1.1 and SRv1.2 respectively for 8xPP7 with 10xSunTag, and 10xPP7 with 12xSunTag. Although mRNA labeled with SRv1, SRv1.1, and SRv1.2 have comparable molecular weights to MS2x128 (Table 3-2), these alternative designs provide flexibility in SunRISER applications to label shorter mRNAs without compromises to mRNA detection.

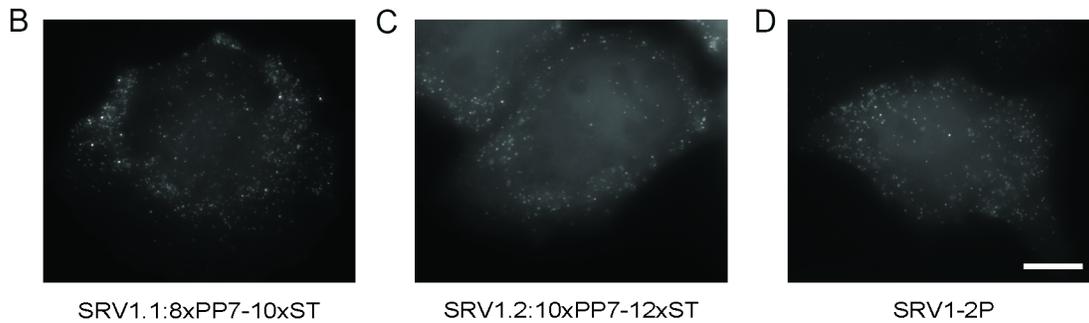
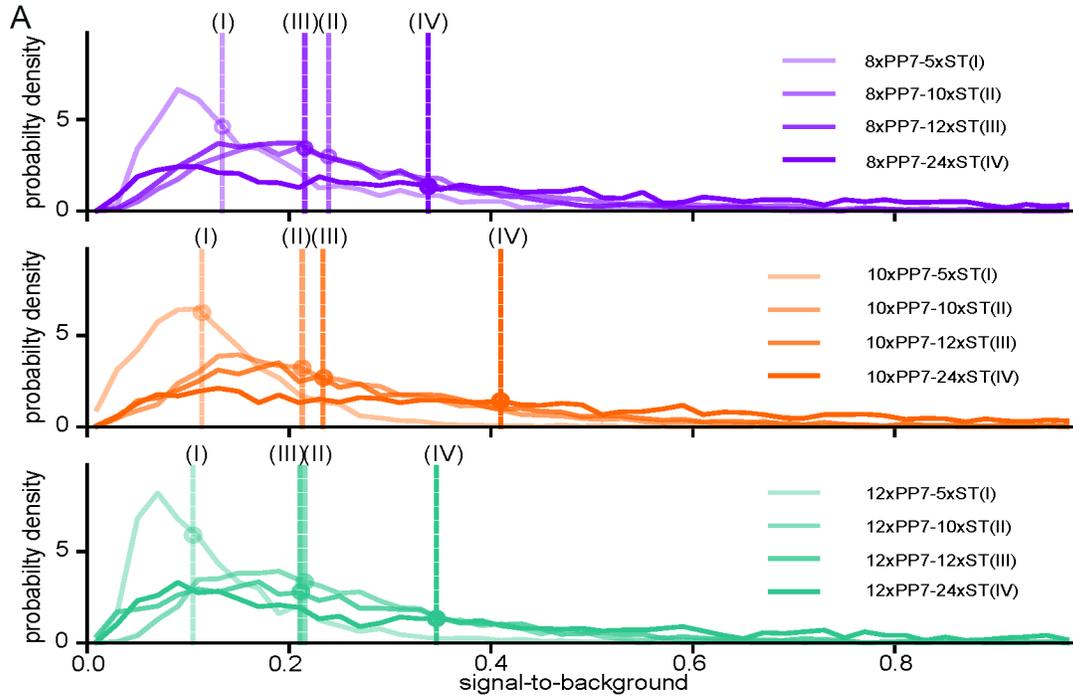
Finally, to further enhance versatility of the SunRISER approach, we considered whether the protein components of SunRISER can be expressed from a single plasmid with multiple promoters. We generated a plasmid with two components to independently regulate expression of GFP-scAB and PCP-5xST, thereby in combination with the detection plasmid reducing the system

to two plasmids (referred to as SRv1-2P; Figure 3-12C). Single mRNAs detected using the SunRISER variant SRv1-2P again showed comparable imaging and detection properties to SRv1 (Figures 3-12 and 3-13). As expected, SRv1.1-2P and SRv1.2-2P also produce results that are indistinguishable from their 3-plasmid counterparts.



**Figure 3-12 Design of SunRISER variants.**

(A) Scatter plot of variance and median for signal-to-background ratio of single mRNAs labelled with different combinations of PP7 stem-loop and SunTag arrays. Vertical and horizontal lines mark the preferred region in the bottom-right quadrant with small inter-spot variance and strong signal. Optimized SunRISER SRv1 is indicated with a star. (B) Boxplots for mRNA numbers detected by different combinations of PP7 stem-loop and SunTag arrays. Grey region marks the first and third quantile for mRNA numbers detected by optimal SunRISER SRv1. Configuration with significantly different mRNA numbers from SRv1 are indicated with red stars (P-value < 0.05, 2 tailed t-test). (C) Diagram of single plasmid encoding 2 protein components (C1) cmv-sfGFP-GB1-scAB and (C2) ubc-nls-PCP-5xSunTag for the SunRISER SRv1-2P variant. (D) Histograms of signal-to-background ratio and mRNA numbers in cells labelled with SunRISER variants SRv1 and SRv1-2P. Adapted from [149].



**Figure 3-13 Quantification and representative images of different configurations for SunRISER.**

(A) Distribution of signal-to-background ratio for different stem-loops paired with different SunTag array lengths. Different color groups indicated different stem-loops and darker shades represent larger SunTag array. Vertical lines mark median values for different combinations. (B) Representative maximum intensity projection images of HeLa cells transfected with *cmv-mCherry-8xpp7*, *phage-nls-pcp-10xSunTag*, *cmv-sfGFP-GB1-scAB* (SR v1.1). (C) Representative maximum intensity projection images of HeLa cells transfected with *cmv-mCherry-10xpp7*, *phage-nls-pcp-12xSunTag*, *cmv-sfGFP-GB1-scAB* (SR v1.2). (D) Representative maximum intensity projection images of HeLa cells transfected with *phage-cmv-*

cfp-24xpp7 and single construct encoding phage-nls-pcp-5xSunTag and cmv-sfGFP-GB1-scAB (SR v1-2P). Scale bar: 10  $\mu$ m. Adapted from [149].

### 3.5 Concluding remarks

Guided by mathematical models, we used synthetic biology to systematically optimize SunRISER design and achieve the optimal molecular configurations that enables mRNA detection at single transcript resolution and maintain high signal intensity and signal-to-background for a long period of time. The same optimization principle is generalizable to orthogonal stem-loop and antibody-epitope pairs and smaller system with shorter stem-loop arrays. We further provided single plasmid version for easier gene delivery. However, the molecular weight of optimal SunRISER is still significantly larger than standard PP7-PCP and could be a potential source for disturbing mRNA metabolism. We continue characterization of SunRISER to investigate if the labeling alters normal mRNA functions.

#### 4.0 Chapter 4. Characterization and application of SunRISER

This chapter is adapted from:

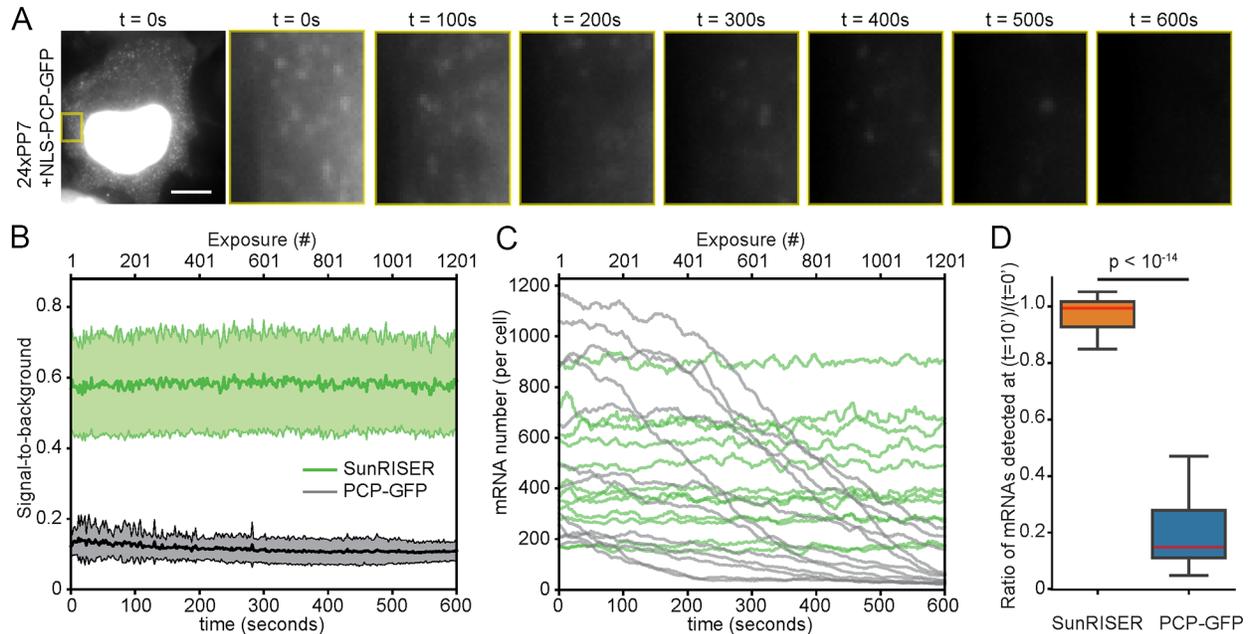
**Y Guo** and REC Lee, Long-term imaging of individual mRNA molecules in living cells, Cell Reports Methods, 2(6), 100226, 2022, of which I was first author, with the following author contributions:

Conceptualization, Y.G. and R.E.C.L.; methodology, Y.G. and R.E.C.L.; investigation, Y.G.; software, Y.G.; formal analysis, Y.G.; writing – original Draft, Y.G. and R.E.C.L.; writing – review & editing, Y.G. and R.E.C.L.; visualization, Y.G. and R.E.C.L.; funding acquisition, R.E.C.L.; supervision, R.E.C.L.

## 4.1 Characterization of SunRISER system

As mRNA is the vital molecule to integrate internal and external environmental cues and dictate the patterns of correct gene expression, perturbation to mRNA metabolism introduced by labeling is always a major concern in applications: whether a specific labeling faithfully reports innate mRNA behavior and whether the alteration is significant and leads to misinterpretation? Stem-loop and its cognate coat-protein as the most widely used technique for imaging mRNA in living cells has been scrutinized for their disturbance effect on mRNA half-life [160] and protein production [161]. So we use general PP7-PCP system as a point of reference to characterize labeling property of SunRISER system.

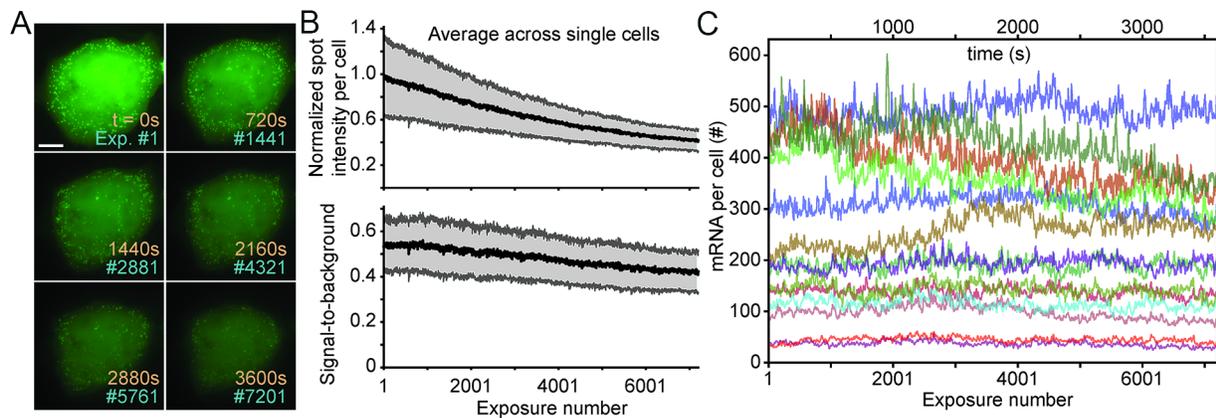
To compare SunRISER directly with widely used approaches, we imaged cells expressing the 24xpp7 detection plasmid with PCP-GFP. Signal-to-background for standard PP7-labeling was near the detection limit throughout, and most cytoplasmic mRNAs fell below detectable signal levels within 10 minutes of imaging (Figure 4-1). By contrast, signal intensity and signal-to-background for SunRISER remained strong and consistent between single cells throughout the 10-minute experiments (Figure 4-1). To characterize photobleaching properties of SunRISER in more challenging imaging conditions, cells were exposed to a rapid time-lapse experiment where 7200 consecutive epifluorescence images were collected within one hour (Figure 4-2). Although signal intensity values decreased to approximately 40% by the end of the experiment, the signal-to-background was reduced only modestly, and mRNAs were robustly detected in single cells throughout (Figure 4-2B and C). Taken together, we conclude that SunRISER is resistant to photobleaching and provides greater robustness for detection of mRNAs when compared with the well-established PP7-PCP systems.



**Figure 4-1 Comparison between mRNAs labeled with SunRISER SRv1 and PP7-PCP-GFP.**

Adapted from [149].

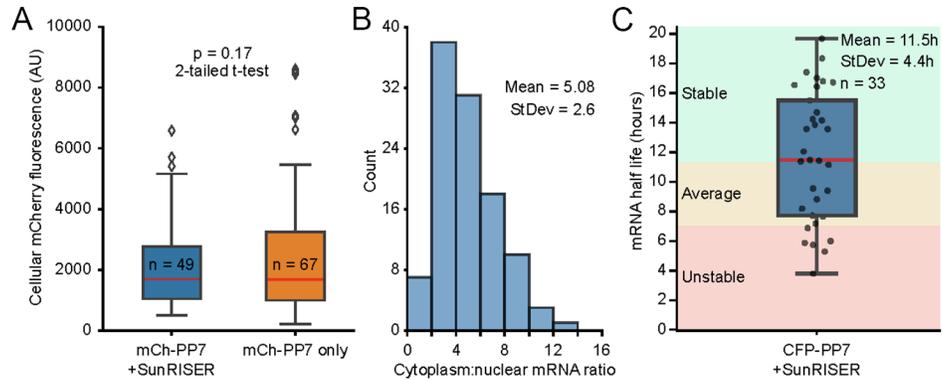
**(A)** maximum intensity projection images of HeLa cells transfected with phage-cmv-cfp-24xPP7 and phage-ubc-nls-ha-pcp-gfp. Cytoplasmic region (yellow box) was zoomed in and shown in time series. **(B)** Representative Time course of signal-to-background values for SunRISER SRv1 (green; n=12 for cell numbers) and PCP-GFP (grey; n=12 for cell numbers) labeled mRNAs. Imaging conditions were selected to minimally facilitate detection of single mRNA spots while at the same time minimizing phototoxicity for the two reporter systems. **(C)** Time course of detected mRNA numbers with SunRISER SRv1 (green) and PCP-GFP (grey). **(D)** Boxplot for ratio of mRNAs detected with SunRISER SRv1(left) and PCP-GFP (right) between t=10min and t=0. The mRNA numbers were calculated as average of 50 frames at specified time points. P-value was calculated using a 2-tailed t-test.



**Figure 4-2 SunRISER is resistant to photobleaching and robustly labels mRNA molecules during prolonged imaging conditions.**

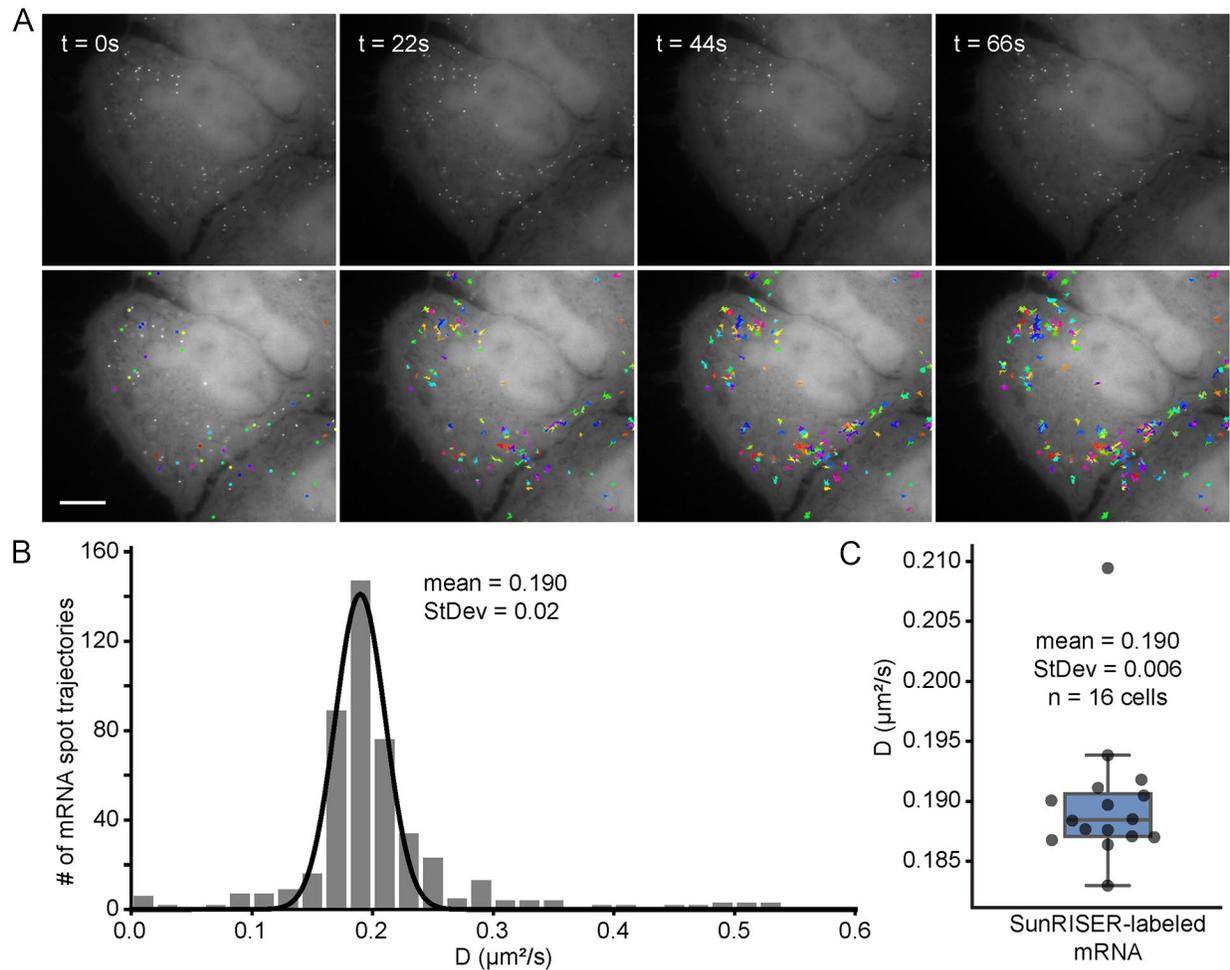
**(A)** Maximum intensity projections of SunRISER SRv1-labeled mRNAs at different time points. 24 hrs after transfection of HeLa cells with detection plasmid phage-cmv-cfp-24xpp7 with ubc-nls-PCP-5xSunTag, and cmv-sfGFP-GB1-scAB (SunRISER), cells were imaged by 60x widefield fluorescence microscopy as 4x frame z-stacks at 2-second interval over a 1-hour duration. Scale bar: 10  $\mu$ m. **(B)** Time course of normalized signal intensity (top) and signal-to-background values (bottom) for spots labeled with SunRISER across different single cells ( $n = 14$ ). Solid line marks the mean signal intensity and mean signal-to-background values averaged among single cells (signal intensity was normalized by the average spot intensity across cells at the first frame of imaging) and shaded area indicates the standard deviation between single cells. **(C)** Time course of mRNA counts for 1-hr fast movies across different cells. The counts were smoothed by a sliding window of 5 frames. Adapted from [149].

Next, we compared functional properties of SunRISER-labeled mRNA with expectations from unlabeled mRNA, as well as previous reports from the literature. Comparing mCherry fluorescence expressed from a detection plasmid in single cells from a transcript with 24xPP7 stem loops, SunRISER labeling did not have significant effects on protein production (Figure 4-3A). SunRISER-labeled mRNA maintained a mean ratio of 5 between nuclear and cytoplasmic compartments (Figure 4-3B), which is consistent with mean value expectations of 3.8 to 6.5 based on RNAseq of subcellular fractions [162], suggesting the subcellular distribution of SunRISER-labeled mRNA is unaltered. Remarkably, SunRISER-labeled mRNA numbers in response to transcription inhibition revealed mRNA half-lives with significant cell-to-cell variability (Figure 4-3C). Single cell mRNA half-lives were consistent with previous results, spanning the divide that separates median mRNA half-lives expected of stable and unstable mRNA molecules [163]. Although this may represent a source of cellular heterogeneity, cells with particularly long mRNA half-lives can also indicate partial escape from transcriptional inhibition through enhanced chemical efflux or other mechanisms that also vary between cells. Finally, extended extensions of widely used RNA tags have the potential to alter diffusion rates, so we also measured diffusive properties of SunTag-labeled mRNA. Since the SunRISER-labeled mRNA complex is larger than 24xPP7-PCP, it is expected to result in lower single-mRNA diffusion rates within the cell. When measured from high-frequency time-lapse images, diffusive motion of SunRISER-labeled mRNA molecules was  $0.19 \mu\text{m}^2/\text{s}$  (Figure 4-4), which is lower yet within the expected range of values ( $0.15\text{-}0.72 \mu\text{m}^2/\text{s}$ ) measured from 24xPP7-PCP-labeled endogenous mRNAs [108]. Taken together, although the SunRISER mRNA-protein complex is bulkier, it does not significantly alter mRNA function beyond expectations from previous analysis of conventional fluorescent mRNA reporters [99].



**Figure 4-3 Characterization of translation, half-life, and cytoplasm-to-nucleus ratio for SunRISER-labeled mRNA.**

(A) Boxplots of mCherry fluorescence intensity in HeLa cells transfected with SunRISER labeled mCherry-24xPP7 (left) and mCherry-24xPP7 only (right) quantified 24 hours after transfection. The p-value (0.17) of a 2-tailed t-test suggests that SunRISER does not significantly alter protein expression from labeled mRNAs. (B) Histogram of cytoplasm:nucleus mRNA ratios for SunRISER labeled mRNAs in HeLa cells transfected with SunRISER labeled CFP-24xPP7 (n=56 for cell numbers). mRNA numbers are counted with dNEMO. (C) Boxplot of mRNA half-lives for SunRISER labeled CFP-24xPP7 mRNAs measured in single cells 24 hours post-transfection. Transcription was inhibited via treatment with 50 $\mu$ M DRB and cells were imaged with 10-minute intervals over a period of 10 hours. To calculate half-lives, single-cell mRNA time courses were fit to exponential functions. Shaded area approximates the half-lives measured for global mRNAs [163, 164]. Adapted from [149].



**Figure 4-4 Diffusion of SunRISER labeled mRNAs.**

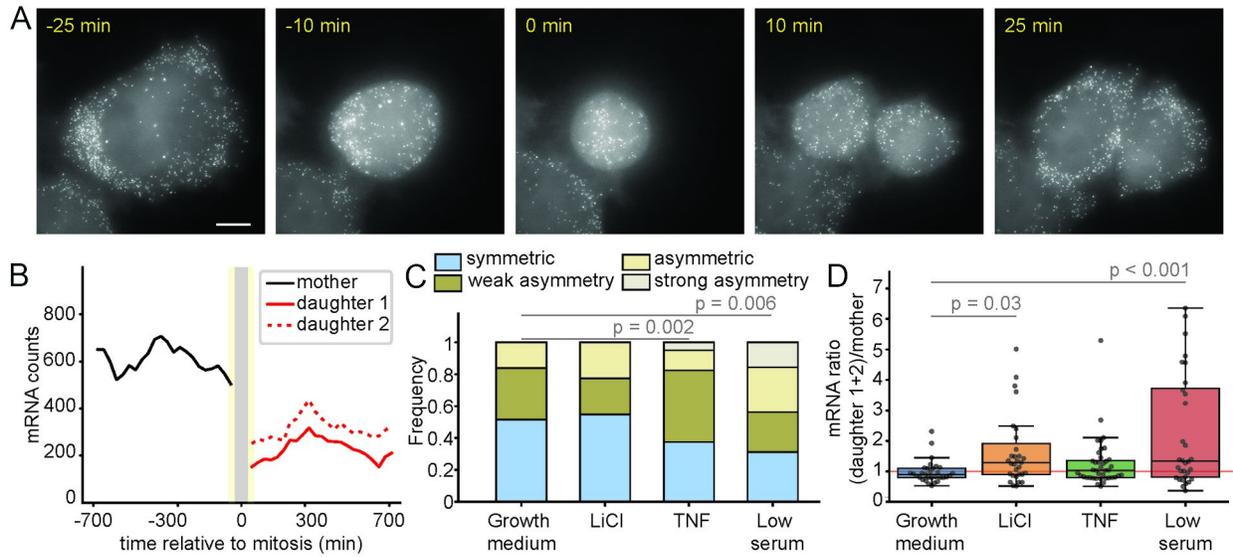
(A) Maximum intensity projection images from a representative high frame rate movie of SunRISER labeled mRNAs. Spots were detected by dNEMO and tracked by u-track [165] to establish single-mRNA time course trajectories. Individual trajectories are distinguished by different colours. Scale bar: 10  $\mu m$ . (B) Representative histogram of diffusion constants calculated from mRNA trajectories within a single cell. Mean and standard deviation are extracted from Gaussian fit ( $n=495$  for spots numbers). (C) Boxplot of mean diffusion constants from 16 cells. Adapted from [149].

## 4.2 Application of SunRISER: Variability of mitotic mRNA inheritance between sister cells

Random partitioning of biomolecules between daughter cells during cell division is a contributing source to non-genetic heterogeneity [4, 166]. The symmetry of mRNA inheritance during mitosis is tightly regulated during embryonic development and tissue homeostasis as the distribution of specific mRNAs controls cell signaling pathways and subsequent cell fate decisions [167-171]. However previous studies on mRNA division have been performed mostly across populations and in fixed samples, therefore lacking the resolution to examine mRNA partitioning in single mother and daughter cells in real time. Since SunRISER is theoretically capable of imaging mRNAs over time scales of cell division, we set out to quantify mRNA partitioning between single mother-daughter and sister cells pairs in various growth conditions.

Enabled by the long-term mRNA imaging capability of SunRISER, we imaged HeLa cells transfected with CMV-promoter driven mRNAs labeled with SunRISER for 24h and quantified mRNA abundance during cell division with and without cellular stress (Figure 4-5A, B). SRv1-labeled mRNAs that encode CFP were used to examine the mitotic partitioning mechanisms for a generic mRNA species that is not subject to particular mitotic regulation mechanisms. The relative difference in mRNA levels between sister cells was used as a metric for symmetry of mRNA partitioning. We observed significant variability between pairs of sister cells in the same culture condition (Figures 4-5 and 4-6). To classify differences between sister cell pairs, we used information criteria and k-means clustering (Figure 4-6A, B) and found 4 clusters that we refer to as ‘Symmetric’, ‘Weak asymmetry’, ‘Asymmetric’, ‘Strong asymmetry’. Cells cultured in the presence of tumor necrosis factor (TNF; 15 ng/mL) or low serum conditions (5% FBS), showed a significant shifts favoring ‘Asymmetric’ with the emergence of rare sister-cell pairs with ‘Strong asymmetry’ (Figures 4-5C and 4-6B). We further compared mRNA ratios between the sum of

daughter cells and their corresponding mother (Figure 4-5D) and observed evidence of significant mitotic transcription (defined as mRNA ratio significantly greater than 1 for the sum of daughter cells divided by the mother cell) induced by cells exposed to lithium chloride (LiCl) as well as low serum conditions (Figures 4-5D and 4-6C, D). Notably, although LiCl promoted escape from silencing of mitotic transcription, it only had modest effects on asymmetric mRNA inheritance between post-mitotic sister cells (Figures 4-5 and 4-6D). For all environmental stresses, distributions of mRNAs allocated to daughter cells shifted significantly away from the binomial distribution observed for cells dividing in the growth medium condition (Figure 4-7A). Although,  $R^2$ -values suggest that partitioning of mRNA between daughter cells is partly explained by distribution of total cellular mass (Figure 4-7B), our data suggest that other factors also contribute to the observed diversification between sister cells.

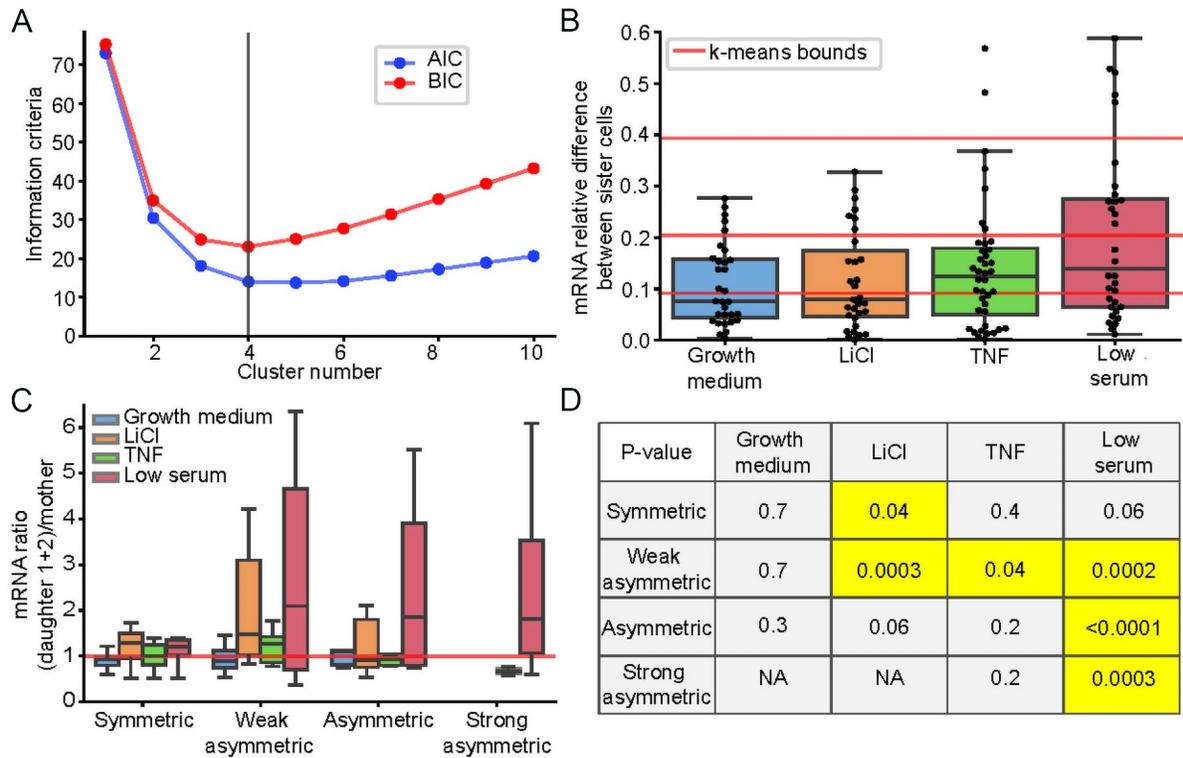


**Figure 4-5 Long-term imaging to monitor mRNA partitioning during cell division.**

(A) Time course for maximum intensity projection images of HeLa cells expressing SunRISER SRv1 in standard growth medium undergoing mitosis. Frame 0 is the last frame where mother cell remains as a single cell with a semi-detached circular appearance. Cells were imaged for 24h with a 10-minute frame rate. (B) Representative trajectories of mRNA counts during mitosis for the cell depicted in panel (A). Solid black line marks mRNA numbers for the mother cell and red lines indicate mRNA numbers for each daughter cell. Grey bar marks the window around Frame 0 during which time mRNA counts are not accurate due to morphological changes and temporary detachment of the dividing cell. Yellow bars mark the 3-frame window used for extraction of mRNA counts for mother and daughter cells in subsequent analysis. (C) Bar graphs for relative difference of mRNA between sister cells during mitosis when cells are cultured in media without or with indicated stress. Classifications are based on information criteria and k-means clustering (see also Figure S9). Yellow colored bars indicate the fraction of cells identified as undergoing asymmetric division with different shades representing the indicated subcategories. Blue colored bars indicate the symmetric division category. P-value was calculated using the Levene test. (D)

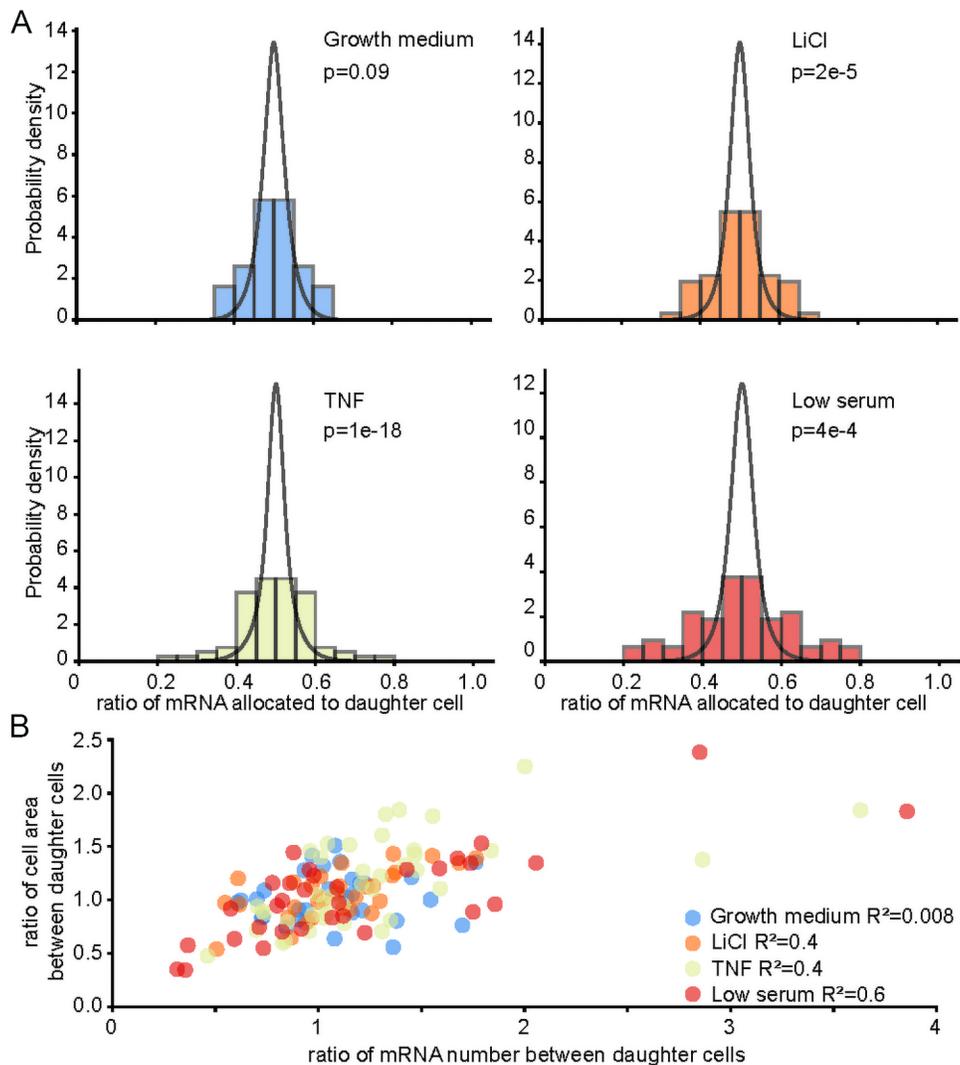
Box plots of ratios for the sum of post-mitotic daughter mRNAs divided by the mRNA number for the pre-mitotic mother cell. Media composition without or with stress is as indicated. Black dots represent individual sister pairs. Red line marks ratio value equal to 1. P-value was calculated using 2-tailed t-tests. Outliers with more than 8-fold increase in transcript numbers are not plotted.

Adapted from [149].



**Figure 4-6 Symmetry and asymmetry of mitotic mRNA partitioning identified by k-means clustering and active transcription during mitosis.**

(A) AIC/BIC calculation for determining the number of clusters ( $k$ ) used for subsequent k-means clustering on mRNA symmetry between post-mitotic sister cells. Vertical line indicates the optimal cluster number as 4 for both analyses. (B) Box plots for relative difference of mRNA between post-mitotic sister cells when cultured in media without or with indicated stress. Black dots represent individual sister pairs. Red lines mark the boundaries of k-means clustering. (C) Box plots for mRNA ratio between the sum of daughters and mother in different symmetry groups identified by k-means clustering on mRNA relative difference. (D) P-value table for t-tests between mRNA ratio between the sum of daughters and mother for each cluster and growth media group shown in panel (C). P-values indicate significant evidence of mitotic transcription in LiCl and low serum conditions. Adapted from [149].



**Figure 4-7 Mitotic mRNA partitioning between daughter cells during stress.**

(A) Histograms for ratios of mRNAs allocated to daughter cells when cultured in media without or with indicated stress. Binomial distributions are overlaid with grey lines. P-values for binomial tests are marked, showing that partitioning deviates significantly from a binomial distribution in all stress conditions. (B) Scatter plots for the ratios of mRNA numbers and the ratios of areas between daughter cells when cultured in media without or with indicated stress.  $R^2$  values are marked. Coefficients of determination suggest that asymmetry of mRNA inheritance is partially, but not completely, explained by variance in cellular mass distribution between daughter cells.

### 4.3 Concluding remarks

The SunRISER system outperforms common PP7-PCP with exceptionally high signal-to-background and resistance to photobleaching for over 7000 consecutive acquisitions. We tested the perturbations of SunRISER to mRNA translation, transport, half-life and diffusion properties and compared the results to reported values in literature. It shows that SunRISER did not significantly alter mRNA functions than PP7-PCP but did reveal heterogeneity in half-life that cannot be measured with ensemble methods. Application of SunRISER to follow mRNA distribution dynamics during mitosis at single mother-daughter pair resolution affords accurate quantification of mitotic inheritance asymmetry in live cells.

## 5.0 Chapter 5 Materials and methods

Further protocol details in applying SunRISER and dNEMO [141] to label and quantify single mRNAs in single cells can be found in Appendix A and based on the paper where I am a co-first author that is in press with Star Protocols:

Y Guo\*, GJ Kowalczyk\*, REC Lee, Label and quantify mRNA molecules in live cell experiments using SunRISER and dNEMO, Star Protocols, In Press.

## 5.1 Plasmid construction

The 24xSunTag-PCP plasmids were constructed with 24xGCN4 repeats flanked by HindIII and BamHI sites and coat protein flanked by BamHI and EcoRI sites in a pcDNA3 vector. Ubc promoter was PCR amplified from phage-ubc-nls-ha-pcp-gfp (Addgene #64539) and inserted between MluI and HindIII sites to make ubc-nxSunTag-PCP. 5xSunTag, 10xSunTag and 12xSunTag variants were generated by PCR amplification of 5xGCN4, 10xGCN4 and 12xGCN4 from pcDNA4TO-5xGCN4\_v4-kif18b-24xPP7 (Addgene #74927), pcDNA4TO-mito-mCherry-10xGCN4\_v4 (Addgene #60914) and 12xMoonTag-12xSunTag-kif18b-24xPP7 (Addgene #128606) and replacing 24xSunTag respectively. SV40NLS, a 57bp NES signal (ATGAACCTGGTGGACCTCCAAAAGAAGCTGGAGGAGCTGGAGCTGGACGAGCAGCAG) or NES from HIV Rev protein and ODC fragment amplified from pEF-24xV4-ODC-24xPP7 (a gift from Dr. Xiaowei Zhuang's lab) were added at the C-terminus of nxSunTag-PCP between EcoRI and XbaI for various modifications. The ubc-nls-PCP-5xSunTag plasmids were created by replacing gfp sequence in phage-ubc-nls-ha-pcp-gfp with 5xGCN4 flanked by BamHI and BsrGI sites. Similarly, ubc-nls-MCP-5xSunTag were obtained by insertion of 5xGCN4 to UbC-NLS-HA-MCP-YFP (Addgene #31230) after digestion of XbaI and BsrGI restriction enzymes.

cmv-sfGFP-GB1-scAB was assembled by inserting sfGFP-GB1 fragment from pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS (Addgene #60904) and scAB fragment from phage UbiC scAB-GFP (Addgene #104998) into a pcDNA3 vector.

phage-cmv-cfp-24xms2 (Addgene #40651) and phage-cmv-cfp-24xpp7 (Addgene #40652) act as reporter mRNA labeled with different stem-loops. 24xMS2V6 stem-loops from pET259-pUC57-24xMS2V6 (Addgene #104391) was amplified using BamHI and SacII sites to

label cfp the same way as other two stem-loops. Phage-cmv-mCherry-24xPP7 was obtained by replacing cfp between AgeI and BamHI with mCherry.

For MoonRISE variant, cmv-sfGFP-GB1-Nb-gp41 was created by replacing scAB with Nb-gp41 from Nb-gp41-Halo (MoonTag-Nb-Halo) (Addgene #128603). Ubc-nls-PCP-12xMoonTag is made by using 12xMoonTag-12xSunTag-kif18b-24xPP7 (Addgene #128606) as backbone and removing sequences after 12xMoonTag and inserting stop codon and inserting ubc-nls-PCP fragment between SpeI and HindIII sites in front of 12xMoonTag.

cmv-mCherry-12xPP7 was constructed by inserting cmv promoter between NotI and BamHI sites into pDZ645 pKAN 1x-mCherry-12xPP7 V4 (Addgene #73173). 8xPP7 and 10xPP7 fragment were amplified from pLH-sgRNA-Sirius-8XPP7 (Addgene #121940) and CMV-10xPP7 oligo-library based (Addgene #158199) respectively and replacing 12xPP7 to create cmv-mCherry-8xPP7 and cmv-mCherry-10xPP7.

Single plasmid encoding SunRISER SRv1 protein components for SRv1-2P was produced by ligating PCR amplified cmv-sfGFP-GB1-scFv-bGH poly(A) into ubc-nls-PCP-5xSunTag backbone.

## **5.2 Live cell imaging and quantification of single mRNA spots with dNEMO**

HeLa cells were plated on 96-well glass bottom plates (Matriplate) at the density of  $8 \times 10^3 \sim 1 \times 10^4$  per well. Transient transfection was performed with Fugene HD (Promega) 24hrs later according to manufacturer's protocol. A mixture of plasmids comprising SunTag system with equal amount (50ng) for each was first made in Opti-MEM then Fugene HD was added and incubated for 15 mins at room temperature. The amount of DNA and Fugene HD can be optimized

accordingly. 24hrs after transfection cells were imaged using a DeltaVision Elite microscope with a 60x objective (1.42 NA; Olympus) and temperature-matched oil in an environmentally controlled chamber (37°C, 5% CO<sub>2</sub>). Z-stacks of 5 images with 1 µm interval were acquired for quantification with dNEMO. Cell segmentation was manually performed in dNEMO and the spot detection parameters are set as default. In comparison of SunRISER and MS2-MCP, stacks of 4 planes with a z-spacing of 0.5 µm were obtained for high frame rate (one 3D stack per 2s). For comparison of promoters, mean fluorescence intensity was measured using ImageJ for a fixed region in cytoplasmic area.

### **5.3 Stress treatments**

Cells were plated on 96-well glass bottom plates (Matriplate) for fixed-cell and live cell imaging experiments. For perturbation of mitosis, cells were cultured in DMEM with 15ng/mL TNF, 10mM LiCl or 5% FBS.

### **5.4 smFISH probes and image acquisition**

Five 3' Cy5 fluorescently labeled DNA oligos [153](ggcaattaggtaccttagg, catatcgtctgctcctttc, gtagtcgacctgcagggag, atatgctctgctggttc, atactgcagccagcgagc) as smFISH probes against PP7 stem-loops were synthesized by Genewiz. Stellaris® Design Ready probes mCherry with Quasar® 670 Dye (VSMF-1031-5) against mCherry CDS were synthesized by Biosearch™ technologies. HeLa cells were plated on 96-well glass bottom plates and transfected

with SunRISE components for 24hrs. Cells were then fixed with 3.7% formaldehyde, washed three times in 1xPBS for 5 mins each and permeabilized in 70% (v/v) EtOH overnight at 4°C. The hybridization was then performed overnight at 37°C with 100 nM probes in 2XSSC with 10% formamide and 10% dextran sulfate. Nuclei were labeled in the wash step after the hybridization. Cells were finally imaged in Glox buffer [172] using a 60X objective on a DeltaVision microscope. Z-stack images of both FITC channel (SunRISER) and Cy5 channel (smFISH) were collected.

### **5.5 Fixed-cell Immunofluorescence**

HeLa cells were plated on 96-well glass bottom plates and transfected with 10xSunTag-PCP or the combination of scAB-GFP and 10xSunTag-PCP for 24hrs. Cells were then fixed with 3.7% formaldehyde for 10 minutes, rinsed three times in 1xPBS for 5 mins each and incubated in 100% methanol for 10 min. Cells were washed three times in PBST (1XPBS 0.1% Tween 20) for 5 mins each and then a primary antibody  $\alpha$ -GCN4 (Absolute Antibody C11L34) diluted 1:100 in 3% BSA PBST was applied and incubated overnight at 4°C. After several washes, cells were incubated with secondary antibody (3% BSA PBST with 1:1000 Alexa594-conjugated anti-mouse IgG antibody) for one hour at room temperature. Nuclei were stained in the wash step after secondary antibody. Cells were finally imaged using a 60X objective on a DeltaVision microscope.

## 5.6 Statistical analysis

Levene test was performed to calculate p-values using raw data from Figure 4-5C, demonstrating significant changes in the variance of the indicated distributions. Binomial test was used to determine if partitioning of mRNA into daughter cells follows binomial distribution. Two-tailed t tests were performed for calculation of p-values for other figures. All analysis is performed with Scipy stats packages in Python.

In Figures 3-12 and 3-13, the same data set is analyzed with the following cell numbers for each condition: 8xPP7-5xST: n=8; 8xPP7-10xST: n=12; 8xPP7-12xST: n=14; 8xPP7-24xST: n=9; 10xPP7-5xST: n=11; 10xPP7-10xST: n=11; 10xPP7-12xST: n=10; 10xPP7-24xST: n=11; 12xPP7-5xST: n=11; 12xPP7-10xST: n=14; 12xPP7-12xST: n=15; 12xPP7-24xST: n=17. In Figures 6C, 6D, S9 and S10, the same data set is analyzed with the following numbers for mother-daughter pairs: growth medium, LiCl, TNF, low serum are 31, 30, 40 and 32, respectively.

Other number of samples are all listed in figure legends or marked in figures.

## 5.7 Half-life and diffusion constant measurement

HeLa cells transfected with SunRISER for 24hrs was treated with 50 $\mu$ M DRB (5,6-Dichlorobenzimidazole 1- $\beta$ -D-ribofuranoside, Sigma-Aldrich) and imaged for 10hrs with frame rate of 10 mins. The time course of mRNA counts was fitted to exponential function  $y = Ae^{-\frac{x}{\tau}}$  to extract mRNA half-life  $\tau_{1/2} = \ln 2 \times \tau$ . For mRNA tracking, fast movie (frame rate of 1.1s) of HeLa cells transfected with SunRISER was taken 24hrs after transfection and mRNAs were identified with dNEMO and tracked with u-track. The diffusion constant is calculated as:

$$D = \frac{e^{\gamma/\alpha}}{6}$$

Where  $\gamma$  and  $\alpha$  are intercept and slope from linear fit to log-log representation of mean square displacement provided by u-track.

### 5.8 mRNA inheritance during mitosis analysis

HeLa cells were transfected with SunRISER for 24hrs and then cultured in DMEM with TNF (15 ng/mL; Peprotech), LiCl (10 mM; Sigma-Aldrich) or 5% FBS and imaged for 24hrs after media replacement with a frame rate of 10 mins. mRNA counts are extracted by dNEMO and mother-daughter pairs are manually assigned. The last frame where mother cell remains as a single cell with a semi-detached circular appearance was set as frame 0. The frames -2, -1, 0, of mother cell and 1, 2, 3, of daughter cells are discarded for mRNA quantification because mRNA counts are not accurate during this period due to morphological changes and temporary detachment of the dividing cell. mRNA counts from frames -5, -4, -3, of mother cell and 4, 5, 6 of daughter cells were averaged to represent mRNA abundance before and after division. mRNA relative difference between sister cells is defined as

$$mRNA \text{ relative difference} = \frac{|daughter1 - daughter2|}{daughter1 + daughter2}$$

It is noted that daughter1 and daughter2 are randomly assigned so we use absolute value of the difference. mRNA ratio between the sum of post-mitotic sister cells and mother cell is defined as  $((daughter1 + daughter2)) / mother$ .

## 6.0 Chapter 6 Discussion

SunTag technology used for fluorescence signal amplification has empowered imaging and quantification of translation, partial story of the central dogma, at single transcript level and also served as a point of reference in dissecting copy numbers of signaling molecules. This highly quantitative tool also imparts an exact control to the fold of signal amplification by varying numbers of SunTag epitope repeats. Another half of the central dogma, transcription, subjected to more tight regulation has been studied extensively at single molecule level but there is still a lack of time-lapse live-cell reporter for resolving single transcripts over extended duration.

The lifespan of mRNA molecules in mammalian cells occurs over timescales of hours to days. During which, mRNAs participate in highly dynamic processes that are tightly regulated in time and space. Bacteriophage-derived stem-loops and FP-tagged coat proteins are current state-of-the-art approaches to detect single mRNA molecules in live cells, however there are still significant limitations of these seminal reporter systems. Furthermore, to take full advantage of typical stem-loop and coat-protein labeling systems requires highly customized microscopy equipment. For example, two-photon fluorescence fluctuation spectroscopy can provide accurate measurement of single mRNAs up to several minutes, but these measurements cannot be sustained over longer durations. By contrast, SunRISER enables long-term investigation of dynamical mechanisms of mRNAs over timescales of at least 24 hours using standard epifluorescence microscopy.

To optimize SunRISER, simulations were used to explore non-linear interrelations between the components of the two-stage reporter system. As a static model, we assume all the binding events in the system have reached equilibrium. The steady state assumption allows

conversion of binding affinity to binding probabilities, which reduces the computational load while preserving key features of the system. Selected binding constants for our simulations were chosen specific to stem-loop, coat-protein, and antibody-epitope pairs used in this study (summarized in Table 3-3). With appropriate modifications to the kinetic parameters, the mathematical model ([https://github.com/reaclelab/SunRISER\\_SupplementalModel](https://github.com/reaclelab/SunRISER_SupplementalModel)) is generalizable to any two-stage molecular amplification reporter and is extensible to higher-order systems.

Direct comparison between SunRISER and PP7-PCP systems showed significant improvements to signal intensity, stability, and signal-to-background ratios. Furthermore, SunRISER minimally perturbs normal mRNA function and is highly resistant to photobleaching. Although our stress-test and long-term experiments concluded after 7200 consecutive images and 24 hours respectively, we expect that these represent lower limits and that mRNA signals will remain detectable over longer repeat-exposure conditions. We also note that several ‘sub-optimal variants’ of the PCP-SunTag component of SunRISER that were tested in our synthetic biology approach showed an exclusively cytoplasmic localization (Table 3-1). Although these sub-optimal variants of the reporter do not identify nuclear mRNAs, they may still have value in certain experimental settings where selective labeling of only cytoplasmic mRNAs is preferred. Finally, SunRISER variants SRv1, SRv1.1, and SRv1.2 offer flexibility to label shorter mRNAs by balancing the reporter’s mRNA:protein composition, reducing sequence perturbations on mRNAs without compromising signal-to-background and detection.

Asymmetric mitotic mRNA inheritance is an essential mechanism to control the maintenance and emergence of specialized cellular phenotypes during development. Previous studies on mRNA partitioning during mitosis typically required cell fixation and chemical synchronization, with mRNA levels assessed via bulk cell measurements that are incapable of

providing dynamical single-cell information. Here we used SunRISER to visualize the mRNA partitioning during mitosis in single cells without chemical perturbations associated with synchronization. We observed significant heterogeneity between sister cells in terms of mRNA partitioning during both TNF stimulation and in low-serum conditions. Remarkably upon serum starvation, a distinct population of cells that exhibit very strong asymmetry in mRNA partitioning arises concomitant with evidence for active mitotic transcription. We surmise that asymmetric segregation of mRNA and rapid divergence via mitotic transcription is a strategy to increase molecular diversity in the cell population. Similar bet-hedging strategies have been observed in yeast and attributed to asymmetric mitotic inheritance of proteins, imparting increased fitness to more diverse populations growing in harsh environments. During stress, such as nutrient limitation, asymmetric partitioning of mRNA in mammalian cells may similarly enhance population-level fitness via diversification of cellular states.

In summary, SunRISER enables unambiguous detection of mRNA molecules in living cells. By using an optimized two-phase design, our reporter system is robust to photobleaching over long term experiments, and the approach is generalizable to other stem-loop and peptide arrays using the accompanying computational tool. We anticipate the approach will facilitate studies of dynamical properties for single mRNAs and biological variability between single-cells, cell types, and eventually in tissues, with applications across biological disciplines.

## 6.1 Limitation of SunRISER

While SunRISER presents a solution for robust labeling of single mRNAs in mammalian cells, there are limitations to be considered. Optimal mRNA labeling with SunRISER requires an expression ratio of protein components using promoters which may be diversely regulated in different cell types and may also vary in different environmental conditions. When applying SunRISER to a new cell line, a stable cell line, or to certain environmental conditions that impact mRNA detection, it may be necessary to test different promoters to ensure the optimal expression ratio is achieved. In our application of SunRISER to study mRNA partitioning during mitosis, a generic transcript was used to examine passive mechanisms of mitotic inheritance of mRNAs. Expectations for mRNA distributions between daughter cells may change significantly for mRNAs associated with specific functions such as mitosis, development, and cell-fate specification. Furthermore, SunRISER-labeled transcripts comprise a large complex that does not significantly alter normal function of mRNAs assessed here, but may still affect other dynamical properties of mRNAs, for example, structural complexes or particular biological processes. When mRNA translation and stability are important aspects of a study, careful controls should be performed to verify there are no specific effects from SunRISER labeling on the particular mRNA species. Therefore, other mRNA labeling methods such as 24xPP7/PCP-GFP may be more appropriate for applications that do not require imaging of cytoplasmic mRNAs, long-term processes, or rapid assemblies such as transcriptional start sites. When SunRISER is used to study mRNA-mRNA interaction or stoichiometry relations, verification via orthogonal approaches will be necessary. It is worth noting that mRNA half-life could be a confounding factor in studying long-term correlations and must be carefully calibrated. Finally, we note that simulations predict SunRISER will perform poorly at extremely high mRNA expression levels. Caution on interpretation should

be taken when mRNA numbers exceed 1000's per cell, or for applications where large numbers of mRNAs are bundled in close proximity.

## 6.2 Future perspectives in application of SunRISER

mRNA is only a small fraction of the genomics information transcribed that encodes proteins [173]. The majority of RNAs in eukaryotic cells is referred to as non-coding RNAs and can be classified into different categories based on their lengths (lncRNA, miRNA), functions (tRNA, rRNA) and localizations (snRNA, snoRNA). SunRISER holds the potential to be applied to non-coding RNAs with shorter stem-loop variant. For example, the standard 24xPP7-PCP cannot be used to label miRNAs as miRNAs are only around 22-nt long [174] and an over 50-fold longer stem-loop appendage would definitely confound the interpretation of observed dynamics. With SunRISER design principle generalizable to shorter stem-loops we propose that 3-5 copies of miRNA labeled with 8xPP7 (SRv.1.1) can be used to study miRNA metabolism. CircRNA [175] is another interesting RNA species that is produced by non-canonical splicing and resistant to common cellular RNA degradation pathways. Similarly, due to the length of 24xPP7, standard PP7-PCP approaches has not been applied to label circRNA. If circRNA can be visualized by shorter SunRISER and followed over its extended lifetime, it will shed light on the degradation mechanism of circRNAs.

Asymmetrical cell division is a fundamental mechanism to produce daughter cells that adopt distinct fates in developmental biology and cancer biology. However, quantitative study of symmetrical properties in mRNA inheritance between mitotic sister cells is limited to biochemical measurements that lack single-cell resolution and dynamical information. With SunRISER, we are

not only able to examine the moment right after division and dissect the direct connection between mother and daughter cells in terms of mRNA counts, but also quantify the potential entanglement between sister cells at varied time scales. It has been proposed [176] that symmetry is established in mitosis via quantum coherence and entanglement among centrioles and mitotic spindles and asymmetry arise from destruction of the communicable entanglement. SunRISER allows following mRNAs dynamics in sister cells for a long time after division so we can determine the correlation in mRNA movements within cells and between sister cells and inspect quantum entanglement in addition to electromagnetic and chemical gradient fields effects. Symmetry in mRNA inheritance is well preserved in normal growth medium and perturbed in stress conditions when a generic gene is studied. When a functional mRNA is tracked, would the symmetry still be observed within its own distribution or intertwined with its related genes or cell-cycle related genes? For example, is it possible that asymmetric inheritance of one mRNA is compensated by another mRNA and the overall mRNA contents partitioned to daughter cells are still even? Are there certain mRNA sequences or structures that serve as the driving force and dictate the proportionate asymmetry of other associated mRNAs? Can the asymmetry of mRNAs expressed upstream predict the symmetrical property of downstream mRNAs? These questions can be answered by examining several related genes at different time points to fully reveal the temporal feature and biological meaning in future applications of SunRISER. With imaging time extended, we can further evaluate whether mRNA asymmetry has functional consequences on the cell lineages, for example in response to subsequent stresses.

## Appendix A

Here we present a copy of the paper listed below currently in press with Star Protocols:

**Y Guo\***, GJ Kowalczyk\*, REC Lee, Label and quantify mRNA molecules in live cell experiments using SunRISER and dNEMO, Star Protocols, In Press.

# Label and quantify mRNA molecules in live cell experiments using SunRISER and dNEMO

**Yue Guo,<sup>1,2,4,6,\*</sup> Gabriel J. Kowalczyk,<sup>1,6</sup> and Robin E. C. Lee<sup>1,3,5,\*\*</sup>**

<sup>1</sup>Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15260, USA

<sup>2</sup>Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15260, USA

<sup>3</sup>Center for Systems Immunology, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15260, USA

<sup>4</sup>Technical contact

<sup>5</sup>Lead contact

<sup>6</sup>These authors contributed equally

\*Correspondence: yug31@pitt.edu

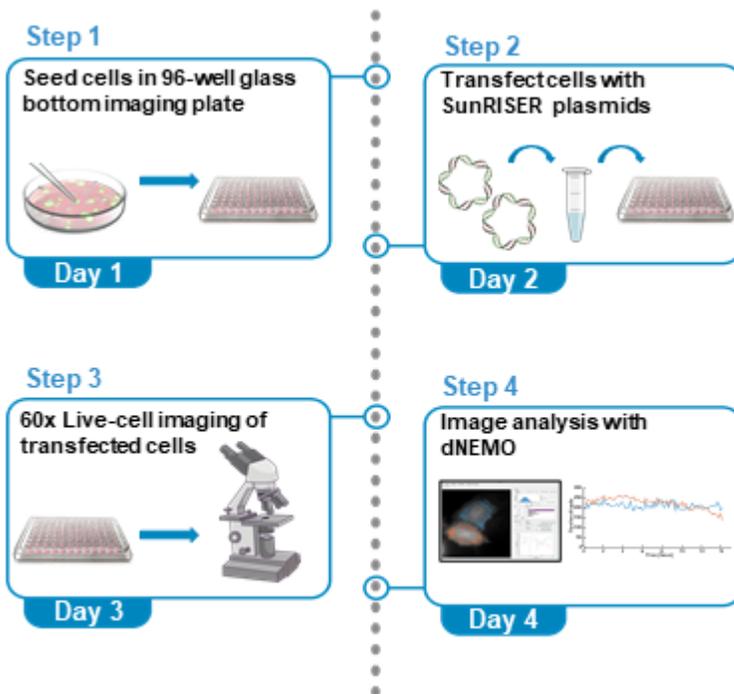
\*\*Correspondence: robinlee@pitt.edu

## Summary

Visualization of mRNA molecules in single cells has revealed core mechanisms of the central dogma as well as sources of cell-to-cell and spatiotemporal heterogeneity. Here we describe a protocol to label, visualize, and quantify mRNA molecules by time-lapse imaging with the capability of resolving mRNA molecules over durations of hours to days. We provide links to mRNA-labeling plasmids as well as free software for a semi-automated image analysis pipeline.

**For complete details on the use and execution of this protocol, please refer to (Guo and Lee, 2022) and (Kowalczyk et al., 2021).**

## Graphical abstract



## Before you begin

The following protocol is for labeling and visualization of single mRNA molecules, followed by image analysis with a semi-automated pipeline. This protocol is described using HeLa cells but is adaptable to other mammalian cells.

## Label a gene of interest with stem-loops in mammalian cells

Single-molecule imaging of mRNA has revealed fundamental properties of mechanisms in the central dogma that lead to variability between single cells. Continuous imaging of mRNA in living cells enables higher spatiotemporal resolution for mRNA shuttling and processing events that occur within the timescales of minutes, and more recently shown in the timescales of hours and days (Cawte et al., 2020; Guo and Lee, 2022; Tantale et al., 2016). To visualize single mRNA molecules for extended durations by wide-field microscopy, we developed a live-cell reporter called SunRISER (Guo and Lee, 2022). SunRISER uses a two-stage labeling approach (Fig. 1). In the first stage, an mRNA of interest is extended in the 3' UTR with a short array of bacteriophage-derived stem-loops (Bertrand et al., 1998; Chao et al., 2008). In the second stage of labeling, individual stem-loop structures are specifically bound by bacterial coat proteins (CP) tagged with a SunTag (Tanenbaum et al., 2014) array of epitopes (CP-SunTag). Fluorescence amplification on an mRNA occurs when CP-SunTag binds to a stem loop, and the SunTag epitope array recruits multiple scFv-GFP molecules (Figs. 1A and 1B).

Previously, we used computational modeling and experiments to establish the optimal SunRISER configuration. An important finding from this work was that a 5:1 protein expression ratio respectively for scFv-GFP and CP-SunTag is necessary for consistent imaging and detection of mRNA molecules. We subsequently demonstrated that a 5:1 ratio can be achieved in HeLa cells by transient transfection using CMV and UBC promoters to drive expression for fluorescent proteins, followed by microscopy and quantitative image analysis (Guo and Lee, 2022).

Although we have had success using this SunRISER configuration to label mRNAs in other human cells, such as A549, this may not always be the case for all cell lines. We therefore suggest for new cell lines, or for stable cell lines, to consider the SunRISER design as a starting point which may require further optimization at the level of promoters to achieve the optimal 5:1 protein expression ratio. We routinely evaluate expression of coding sequences (CDS) for fluorescent proteins for different promoters, using microscopy or flow cytometry to establish the average fluorescence among a population of cells.

Another important consideration is the choice of which RNA species to visualize for a SunRISER-labeling experiment. We evaluated SunRISER using the CDS for mCherry to represent a generic template that does not undergo particular regulation. In general, we expect that many SunRISER-labeling experiments will opt to switch the CDS template to other sequences (described below). See (Guo and Lee, 2022) for complete details on the modeling, optimization, and experimental validation of the SunRISER labeling system, as well as description of different SunRISER variants. Component plasmids for different versions of SunRISER (Fig. 1C and Table 1) are available on Addgene (see the Key Resource Table).

### Timing: 2 weeks

Clone the gene of interest (GOI) into a reporter expression vector suitable for the specific application. Among the SunRISER plasmid toolkit, we provide 2 different stem-loop array lengths of PP7 stem-loops (8xPP7, 10xPP7), and plasmids for 24xPP7 are also available (Addgene #40652). In these plasmids, the CDS for the mCherry reporter gene is flanked by standard restriction enzymes to facilitate replacement with another GOI via typical molecular cloning methods. Although the plasmids have repeat sequences which can lead to technical difficulties, in our experience we have had no issues inserting GOIs and plasmid amplification in typical *E. coli* strains such as DH5-alpha is routinely successful. See (Guo and Lee, 2022) or Addgene for further information and plasmid maps. The resulting plasmid containing the GOI with stem loop extensions is referred to as the 'detection plasmid'.

## Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
---------------------	--------	------------

Chemicals, peptides, and recombinant proteins		
Fugene HD	Promega	Cat # E2311
Opti-MEM, Reduced Serum Medium	Thermo Fisher Scientific	Cat # 31985062
DMEM	Corning	Cat # 10-017-CV
FBS	Corning	Cat # 35-010-CV
Penicillin-Streptomycin (10,000 U/mL)	Thermo Fisher Scientific	Cat # 15140122
L-Glutamine (200 mM)	Thermo Fisher Scientific	Cat # 25030081
FluoroBrite™ DMEM	Thermo Fisher Scientific	Cat # A1896701
Experimental models: Cell lines		
Human: HeLa cell	ATCC	RRID: CVCL_0030
Recombinant DNA		
cmv-sfgfp-gb1-scAB	(Guo and Lee, 2022)	Addgene #185794
cmv-mCherry-8xPP7 (detection plasmid)	(Guo and Lee, 2022)	Addgene #185795
cmv-mCherry-10xPP7 (detection plasmid)	(Guo and Lee, 2022)	Addgene #185796
ubc-nls-pcp-5xSunTag (SRv.1)	(Guo and Lee, 2022)	Addgene #185797
ubc-nls-pcp-10xSunTag (SRv.1.1)	(Guo and Lee, 2022)	Addgene #185798
ubc-nls-pcp-12xSunTag (SRv.1.2)	(Guo and Lee, 2022)	Addgene #185799
cmv-sfgfp-gb1-scAB-ubc-nls-pcp-5xSunTag (SRv.1-2P)	(Guo and Lee, 2022)	Addgene #185800
cmv-sfgfp-gb1-scAB-ubc-nls-pcp-10xSunTag (SRv.1.1-2P)	(Guo and Lee, 2022)	Addgene #185801
cmv-sfgfp-gb1-scAB-ubc-nls-pcp-12xSunTag (SRv.1.2-2P)	(Guo and Lee, 2022)	Addgene #185802
phage-cmv-cfp-24xpp7 (detection plasmid)	(Wu et al., 2012)	Addgene #40652
5-alpha Competent <i>E. coli</i> (high efficiency)	New England Biolabs (NEB)	Cat # C2987H
Software and algorithms		
ImageJ	(Schneider et al., 2012)	<a href="https://imagej.nih.gov/ij/">https://imagej.nih.gov/ij/</a>
dNEMO	(Kowalczyk et al., 2021)	<a href="https://github.com/reacleelab">https://github.com/reacleelab</a>
Cellpose	(Stringer et al., 2021)	<a href="https://github.com/MouseLand/cellpose">https://github.com/MouseLand/cellpose</a>
Python	Python Software Foundation	<a href="https://www.python.org">https://www.python.org</a>

Digital resources associated with this protocol	This paper	DOI: 10.17632/8j4x6dj2f7.1
Other		
DeltaVision Elite	GE	N/A
96-well glass bottom plate	Matriplate	Cat # MGB096-1-2-LG-L

## Step-by-step method details

### SunRISER plasmids delivery into HeLa cells to label single mRNAs

Timing: 4 days

This step allows delivery of GOI labeled with stem-loops and SunRISER labeling component plasmids via transient transfection.

1. Choose the appropriate SunRISER setup for your application
  - a. Choose the variant of SunRISER that favors either a shorter or longer stem-loop extension (Fig.1 and Table 1) and select the appropriate detection plasmid available from Addgene (SRv.1: 24xPP7; SRv.1.1: 8xPP7; SRv.1.2: 10xPP7) or use the user-generated detection plasmid as described in the ‘Before you begin’ section above.

**Table 1. SunRISER plasmid variants.**

SunRISER system	Detection plasmid	pcp-nxSunTag	scFv-GFP
Variation 1 – SRv.1	Phage-cmv-CFP-24xpp7 (Addgene #40652)	ubc-nls-pcp-5xSunTag (Addgene #185797)	cmv-sfgfp-gb1-scAB (Addgene #185794)
Variation 2 – SRv.1-2P	Phage-cmv-CFP-24xpp7 (Addgene #40652)	cmv-sfgfp-gb1-scAB-ubc-nls-pcp-5xSunTag (Addgene #185800)	
Variation 3 – SRv.1.1	cmv-mCherry-8xPP7 (Addgene #185795)	ubc-nls-pcp-10xSunTag (Addgene #185798)	cmv-sfgfp-gb1-scAB (Addgene #185794)
Variation 4 – SRv.1.1-2P	cmv-mCherry-8xPP7 (Addgene #185795)	cmv-sfgfp-gb1-scAB-ubc-nls-pcp-10xSunTag (Addgene #185801)	
Variation 5 - SRv.1.2	cmv-mCherry-10xPP7 (Addgene #185796)	ubc-nls-pcp-12xSunTag (Addgene #185799)	cmv-sfgfp-gb1-scAB (Addgene #185794)
Variation 6 - SRv.1.2-2P	cmv-mCherry-10xPP7 (Addgene #185796)	cmv-sfgfp-gb1-scAB-ubc-nls-pcp-12xSunTag (Addgene #185802)	

**Note:** Although SRv.1.1 and SRv.1.2 extend the target mRNA with a smaller stem-loop array than SRv.1, the associated protein component (consisting of pcp-nxSunTag and

scFv-GFP proteins, see Fig. 1) of the labeled-mRNA complexes for SRv.1.1 and SRv.1.2 are larger than for SRv.1. Because of this complementarity, SRv.1 and SRv.1.2 are similar in overall molecular weight. The overall molecular weight of labeled mRNA complexes for SRv.1.1 is approximately 35% smaller than for SRv.1 and SRv.1.2, as described in (Guo and Lee, 2022). For all SunRISER variants, the mRNA:protein composition of the labeled mRNA is balanced for optimal signal-to-background. The user can therefore select a SunRISER variant that works best with their application. For example, extending shorter mRNAs with a smaller stem-loop array reduces sequence perturbations, which may be favorable, but may not necessarily lead to the best mRNA detection. The choice of SunRISER variant should ultimately weigh the particular application and experimental goals against empirical imaging results.

- b. Based on the stem-loop lengths, choose the corresponding SunTag array lengths. (SRv.1: 24xPP7-5xSunTag; SRv.1.1: 8xPP7-10xSunTag; SR.v1.2: 10xPP7-12xSunTag).
  - c. Choose the plasmid version suitable for your system. In SunRISER, we provide both a single plasmid encoding both scFv-GFP and pcp-nxSunTag and a version where the two constructs are on separate plasmids (see Fig.1 and Table 1). Separate plasmids can be useful in determining proper expression ratios are achieved as well as to alleviate any unexpected complications that could arise from a double-expression plasmid.
  - d. Calculate the amount of each component based on the selected version of SunRISER. For three-plasmids, equal molar amounts of each plasmid (1:1:1) is optimal. Since the three plasmids as supplied have comparable size, this also works out to a 1:1:1 weight ratio. Similarly, a 1:1 molar ratio is optimal for the two-plasmid version. For the 2-plasmid version as supplied, note that a 1:1 molar ratio equates to a 1.5:1 weight ratio because the double expression plasmid as supplied is approximately 1.5x the size of the detection plasmid. This ratio may change depending on the size of the GOI.
2. Seed HeLa cells in 96-well imaging plates.
    - a. Seed  $1 \times 10^4$  HeLa cells per well of 96-well glass bottom imaging plates (Matriplate; .17mm Flat Clear Glass Bottom) in 300 $\mu$ L of DMEM (supplemented with 10% FBS, 1% streptomycin/penicillin, and 1% L-glutamine)

**Note:** While culture conditions for HeLa cells as described in this protocol do not require coating (fibronectin, collagen, etc.), other cell lines may. We do not expect any compatibility issues in imaging SunRISER with cell lines using fibronectin, poly-L and poly-D lysine. Other coatings such as collagen should be tested to ensure they do not have autofluorescence properties that interfere with fluorescence imaging.
    - b. Incubate cells at 37°C in 5% CO<sub>2</sub> incubator overnight (18-24 hours) to allow cells to recover and adhere to the plate at approximately 70% confluency.
  3. Transfect cells with SunRISER plasmids using FuGENE<sup>®</sup> HD following [manufacturer's instructions](#).
    - a. Warm FuGENE<sup>®</sup> HD Transfection Reagent to room temperature.

- b. Mix DNA and FuGENE<sup>®</sup> HD Transfection Reagent according to manufacturer's instructions. For this protocol, we mixed 150 ng total plasmid with 0.45  $\mu$ L FuGENE<sup>®</sup> HD Transfection Reagent in 10  $\mu$ L Opti-MEM<sup>™</sup> for each well. [See Troubleshooting Problem 1]
- c. Incubate the mixture for 10-15 minutes at room temperature.
- d. Add mixture into each well and mix by pipetting.
- e. Return cells to incubators for 24-48 hours. Post-transfection media is supplemented with 10% FBS, 1% streptomycin/penicillin, and 1% L-glutamine. The cells should be approximately 85% confluent post-transfection.

**Note:** Each step can be optimized with the [instructions](#) from the FuGENE<sup>®</sup> HD transfection protocol. Although we have only used FuGENE<sup>®</sup> HD for our SunRISER-labeling experiments, we expect other transfection reagents and techniques (such as electroporation) will lead to equivalent results.

## Imaging of SunRISER-labeled mRNAs

### Timing: 2 days

This step allows imaging of SunRISER-labeled mRNAs for up to 24 hours. For time-lapse live cell imaging we used a DeltaVision Elite microscope (GE Healthcare) equipped with a 60x NA1.42 oil-immersion objective, sCMOS camera, solid-state illumination module, and an environmental controlled chamber (37°C, 5% CO<sub>2</sub>). A minimum of 60x magnification is required to resolve single mRNA puncta as diffraction-limited objects (Fig. 2, see also (Guo and Lee, 2022)).

**Note:** We expect any epifluorescence microscope equipped with an environment control chamber can be used for imaging SunRISER-labeled mRNAs. Other more advanced microscopes (e.g., confocal, light sheet, and many others) are likely to produce SunRISER images with even greater signal-to-background. We have successfully imaged SunRISER in 24-hour experiments and expect that longer duration experiments may also be possible.

4. Before beginning, replace complete growth medium with FluoroBrite medium (supplemented with 10% FBS, 1% streptomycin/penicillin, and 1% L-glutamine).

**Note:** FluoroBrite or any phenol-red free cell culture medium is recommended to increase the signal-to-background of fluorescence images.

5. Place the imaging plate in the microscope incubation chamber to allow the whole system to equilibrate for at least 30 mins before the experiment starts.
6. Set up appropriate imaging conditions for live cell imaging. For 24-hour imaging, we select FITC as Ex and Em filters and set the exposure time to 0.005 sec and ND filter to 10%. Images were taken as Z-stacks with 1  $\mu$ m slices spacing and total thickness 5  $\mu$ m every 10 mins for 24 hours. Data acquisition was performed in 1024  $\times$  1024-pixel format.

**Note:** Multi-channel imaging using spectrally compatible filter sets can be used to select for successfully transfected cells based on expression of mCherry or CFP if these are the GOIs expressed from the detection plasmid.

**Note:** For a typical widefield microscope equipped with a NA1.4 60x oil objective, 0.3  $\mu\text{m}$  z-spacing will be approximately the ideal Nyquist sampling, where each diffraction-limited object will appear in at least two to three consecutive z-slices. Although the z-spacings between 0.5 and 1  $\mu\text{m}$  will not oversample as effectively as Nyquist conditions, it allows coverage over wider axial range with fewer exposures at each time point. In our experience these settings lead to quantitatively similar results for mRNA numbers, while reducing the effects of photobleaching and phototoxicity in long-term experiments. For particular applications, ideal combinations of z-spacing and z-slice numbers can be chosen to favor oversampling for increased spatial resolution in the z-axis, or increased depth, or reduced molecular and cellular strain as appropriate to the usage case. Experiments demonstrating resistance to photobleaching with these settings can be found in (Guo and Lee, 2022) .

7. Select cells expressing SunRISER-labeled mRNA as diffraction-limited spots and mark the region of interest for live-cell imaging experiments.

**Note:** Imaging conditions are considerably variable between different microscope setups. SunRISER-labeled mRNAs appear as diffraction-limited spots (Fig. 2A, B), and the raw pixel intensity range for the spots detected under the imaging conditions described in this protocol have an average intensity of 673 +/- 189 units from a 16-bit sCMOS detector when pixel values are examined in image analysis software like ImageJ (Schneider et al., 2012). For cells expressing lower scFv-GFP abundance, mRNA spots can be difficult to distinguish with a poor signal-to-background ratio and can suffer from rapid photobleaching if the excitation illumination is too intense. [See Troubleshooting Problem 2]

**Note:** Transient transfection can lead to subpopulations of cells with GFP expression that is significantly higher than average. We typically avoid imaging and analysis of cells expressing too much GFP that show very bright background fluorescent intensity, and cells with spots that are larger than the diffraction limit Fig. 2C). [See Troubleshooting Problem 3]

8. When possible, use Ultimate Focus or a related auto-focusing mechanism to prevent the loss of the imaging planes during long-term imaging experiments. [See Troubleshooting Problem 4]
9. Begin time-lapse imaging experiment. The resulting movie file can be opened with ImageJ and quantified with dNEMO as described below. The included file 'SunRISER\_SAMPLE\_MOV.tif' (DOI: 10.17632/8j4x6dj2f7.1; see also, Fig. 3A) is representative of a typical imaging experiment result using HeLa cells transfected with stock 24xPP7 and SRv.1-2P plasmids as described here.

## Image analysis with dNEMO and Cellpose

**Timing:** ~20 minutes (for sample SunRISER movie with 145 time points each with 4 z-slices)

**Note:** The timing for this step can increase depending on the number of timepoints in a given time-lapse series, the number of cells in the given image, and the hardware running both dNEMO and Cellpose. The reported time is given for the sample SunRISER image provided alongside the software on a 2015 MacBook Pro laptop (16 GB RAM, 2.5 GHz processor).

This section describes the semi-automated image analysis which identifies individual mRNA transcripts, extracts the fluorescence intensity from each transcript, and assigns mRNAs to individual cells to create single-molecule single-cell datasets over time-lapse images (Fig. 3). We use the spot detection and quantification tool dNEMO (detecting-NEMO) which is optimized for rapid and accurate detection of punctate structures (spots) in time-lapse fluorescence microscopy images. While dNEMO contains dedicated tools for the manual segmentation of individual cells, the most recent version of the software interfaces with Cellpose (Stringer *et al.*, 2021), a generalist algorithm for automated cell and nucleus segmentation. Cellpose automatically segments multiple cells in each input image and typically performs better when cells are sub-confluent. With HeLa cells, we typically have less than 5 cells per 60x image with a 1024x1024 sCMOS detector. Individual cells are reconstituted as single-cell trajectories when input into dNEMO. In this protocol, the combination of dNEMO and Cellpose is used for automated spot detection, cell segmentation, and cell tracking. Sample images acquired using SunRISER and the corresponding results files from both dNEMO and Cellpose are provided with this protocol (DOI: 10.17632/8j4x6dj2f7.1). Also included for users is an additional sample image of cells labeled with smFISH probes and corresponding dNEMO/Cellpose results files.

**Note:** There are standalone executables of the dNEMO software available with this protocol (DOI: 10.17632/8j4x6dj2f7.1) which do not require MATLAB to be installed. Use of Cellpose and dNEMO outside of the standalone executables requires installation of Python 3, MATLAB, and the OME bioformats package. Detailed installation instructions can be found in the dNEMO documentation on <https://github.com/recleelab>.

**Optional:** If Cellpose is not installed, dNEMO implements a manual cell segmentation tool which can be used to segment cells within the dNEMO interface.

10. Open the time-lapse image in dNEMO.
  - a. To open the standalone dNEMO executable, navigate to the dNEMO application within the executable folder and double-click the application icon.

**Alternatives:** If using the MATLAB script package, type the following into MATLAB's command window:

```
>addpath(fullfile(cd, dNEMO_MATLAB_scripts))  
>RUN_ME
```

- b. With the interface open, navigate to File > Load Images. Select the image to be analyzed in the subsequent file selection pop-up window. The image 'SunRISER\_SAMPLE\_MOV.tif' is provided with the software for this protocol (DOI: 10.17632/8j4x6dj2f7.1) and can be opened and analyzed using the settings and steps

described here. This image (Figs. 3 and 4A) is also depicted in subsequent figures as an example for detection of SunRISER-labeled mRNAs within the dNEMO interface.

**Note:** Image formats supported by BioFormats (Linkert et al., 2010) can be opened by dNEMO. Upper limit on image size may vary with available system memory.

11. Run spot detection on the currently displayed image.
  - a. Select 'Test Detect (Full)' in the upper right 'Spot Filter' panel. This will run the spot detection algorithm over the currently displayed image (Fig. 4A).
  - b. Confirm visually that all spots are being detected by toggling the 'Display All Signals in Current Frame' toggle in the 'Display' panel.
    - i. Adjust the value in the 'Wavelet Threshold' box (Fig. 4A) and click 'Test Detect (Full)' (Fig. 4A), both in the 'Spot Filter' panel, to re-run spot detection with an updated threshold value. The higher this value is, the fewer objects will be considered above the watershed threshold and thus detected. For images shown detecting SunRISER-labeled mRNAs in this protocol, the 'Wavelet Threshold' value was set to 2.25.

**CRITICAL:** Properly detecting spots in an image requires examining the detected spots and adjusting the parameters as needed. Two additional parameters for spot detection that affect the generation of the wavelet map and cross-referencing the resulting wavelet maps of a 3D image stack can be found in Settings > Signal Parameters (Fig. 4B). The 'Frame Limit' parameter indicates the number of consecutive slices in a 3D stack a spot must be detected in to be considered valid. In order to detect spots which only appear in one slice, for example, this value would need to be set to 1. The 'Wavelet Level' parameter dictates the level of the wavelet transform to use when generating the wavelet map to detect spots. The higher this value is, the larger the objects detected will be in the resulting wavelet map. In detecting the SunRISER-labeled mRNAs for this protocol, the 'Wavelet Level' was set to 2 and the 'Frame Limit' was set to 1. All other settings were kept at the default values. For complete details on the implementation of the wavelet transform and watershed segmentation operations in dNEMO, see (Kowalczyk *et al.*, 2021). [See Troubleshooting Problem 5]

- c. Once satisfied that the spots are being accurately detected, run spot detection with the current detection parameters over all images in the time-series by clicking 'Create Keyframe' in the upper right 'Spot Filter' panel (Fig. 4A). The Keyframes box (Fig. 4C) will update with information for detected spots and the associated parameters.
12. Run Cellpose to segment single cells over the time-lapse image sequence.
  - a. Initiate Cellpose from dNEMO by navigating to Cell Masks > Run Cellpose (Fig. 5A). This will startup Cellpose and run on the movie currently loaded into dNEMO. [See Troubleshooting Problem 6]
  - b. Upon completion, dNEMO will display a prompt for confirming the imported mask. Click 'OK' to confirm the mask import.

**Note:** Cellpose does not provide any tracking of the segmented cells. Cells are tracked over time using several parameters found in dNEMO by navigating to Cell Masks > Adjust Import Settings. For more information on how these parameters function see documentation of dNEMO at <https://github.com/recleelab>.

**Alternatives:** If Cellpose is not installed on your system, manual segmentation is an option implemented in dNEMO. Similarly, cell masks generated in other applications can be imported using the 'Cell Masks' dropdown menu provided the masks are in a compatible matrix format (TIFF or excel/csv spreadsheet). Manual segmentation uses a process called keyframing where user-defined cell boundaries are propagated across frames of a time-lapse image. Briefly, the 'Add Cell' button of the 'Cells' panel will create an interactive polygon drawing tool over the current image (Fig. 5B). Clicking on the image will begin the manual segmentation process and completing the polygon will create a new cell. The polygon for each cell can be left alone or modified as needed to create subsequent keyframes to refine changes to cell morphology or position at later time points. See (Kowalczyk *et al.*, 2021) for full description of keyframing for cell segmentation and additional spot detection parameters.

- c. Cells can be adjusted using both the 'Cells' and 'Keyframes' panels in dNEMO.
  - i. The slider along the bottom of the image can be used to navigate through the frames of a time-lapse image.
  - ii. Select a cell using the 'Cell Selection' dropdown menu in the 'Cells' panel or clicking on a cell in the 'Keyframes' panel (Fig. 5C).
  - iii. Segmentations can be adjusted for the current frame by clicking the 'Modify Cell' button in the 'Cells' panel.
  - iv. Multiple operations to adjust segmentations over time and deleting a cell can be accessed by right-clicking on a cell in the 'Keyframes' panel and selecting an option from the pop-up menu (Fig. 5D). In conjunction these can be used to modify segmentations or delete inaccurate segmentations. For complete details on dNEMO's keyframing functions, see (Kowalczyk *et al.*, 2021) and <https://github.com/recleelab>. [See Troubleshooting Problem 7]
13. Further curate data using keyframing tools and manual exclusion tools.
  - a. Spots can be curated in an automated fashion by assigning keyframes for detected spots' physical features. For example, using the histogram display axis in the 'Spot Filter', detected transcripts can be limited to those which have maximum intensities greater than or equal to 0.0075 by selecting 'Max' from the dropdown menu below the axis and typing 0.0075 into the 'Min' value to the right of the axis (Fig. 6A, B).
  - b. Click 'Create Keyframe' in the spot filter panel to create a new keyframe for this feature. This creates a parameter for the upper and lower bounds of fluorescence intensity for spots deemed as acceptable. These bounds are propagated over the entire movie and can be adjusted by creating additional keyframes at different time-points. The bounds can also be deleted by selecting and right-clicking a given parameter in the 'Manual Curation' section of the 'Keyframes' panel (Fig. 6C).

- c. Detected spots which are determined erroneous (e.g., lysosomal accumulation of fluorescent molecules) or oversegmented (Fig. 3C) can be curated manually using the manual removal tool. [See Troubleshooting Problem 8]
  - i. Navigate to the 'Manual Removal' panel and click 'Remove Signals' (Fig. 6C).
  - ii. A crosshair will replace the mouse pointer icon when hovering over the main image axis in dNEMO. Clicking on detected spots in the image will remove them. Clicking on them again will undo the removal. Clicking and dragging the cursor will create a box allowing removal of numerous detected spots.
  - iii. Click 'Update Removal Keyframe' to save the manual exclusions performed.
  - iv. Click 'Stop Removing Signals' in the Manual Removal Panel to terminate the manual exclusion process.
14. Save data collected in dNEMO to mat-file and excel spreadsheet.
  - a. Navigate to File > Save to output results. This will result in a number of output files which are detailed later in this protocol and within dNEMO's documentation on <https://github.com/reaclelab>.

## Expected outcomes

For over two decades, techniques such as single molecule fluorescent in situ hybridization (smFISH) have been powerful tools to resolve single molecules of RNA (Femino et al., 1998; Raj and van Oudenaarden, 2008). Studies using smFISH have revealed mechanisms of gene expression and numerous consequences of cell-to-cell heterogeneity but are generally limited to single timepoints because hybridization requires cell fixation. As a live-cell reporter, transfection of HeLa cells with SunRISER components enables high-intensity and photostable labeling of individual mRNA molecules that can be imaged by wide-field fluorescence microscopy for many hours (Fig. 3). When analyzed with dNEMO, the number and intensity of mRNA spots can be quantified over time in each frame of the time-lapse image (Figs. 3C and 3D). After the mRNA spots are detected across frames of the time-lapse images, cells can be manually segmented, automatically segmented using the Cellpose software, or a combination thereof to generate single-cell time-lapse RNA datasets. These results are written to an excel file and several mat-files, one ('full\_results' mat-file) which can be reloaded into dNEMO for additional analysis at later time points (Fig. 3D). An optional AVI file depicting circles around transcripts detected within the segmented cells as displayed in the dNEMO interface can also be created by the user by navigating to File > Save as AVI.

## Quantification and statistical analysis

Additionally included in the MATLAB version of the software package for this protocol is an interface for batch processing a folder of image files to automate spot detection and cell segmentation. To run, ensure the 'batch\_processing' folder is both within the dNEMO directory and on the current path in MATLAB and type the following into the command window:

```
>RUN_ME_BATCH
```

On the lefthand side of the interface, select the input directory containing image data and the output directory to save results. The center panel contains the parameters for detection of spots, keyframing spots' features, Cellpose segmentation import parameters, and additional settings. These can be interacted and set manually, or a text file containing preset parameters can be loaded into the workspace by clicking the 'Browse' button on the top of the center panel. At the bottom of the center panel, you can select which of the 2 major operations (dNEMO and/or Cellpose) should be propagated over the selected input image directory. The right panel of the interface contains a log window which records what operations are happening on which input files. When a valid input directory, output directory, and operation settings are selected, click 'Confirm Valid Input Arguments' to lock in the settings and 'Start' to run operations over the selected input directory. For full details on the arguments input into the batch processing tool, please see <https://github.com/recleelab>.

## Limitations

When adapting SunRISER for cell lines other than HeLa, the promoters used in HeLa cells may require optimization to achieve the most effective expression levels for SunRISER labeling. It was previously shown in (Guo and Lee, 2022) that SunRISER is capable of resolving small numbers of transcripts in cells, and while SunRISER is optimized for expression on the order of hundreds of transcripts, caution in interpretation of results should be observed for mRNA numbers exceeding the order of thousands per cell. We suggest using SunRISER as a starting point and further calibrate the strengths of promoters in your cell lines to reach desirable labeling. The expression ratio of protein components in different cell types may similarly need to be adjusted through the use of different promoters to reach the 5:1 protein expression ratio for optimal mRNA labeling. As with other live-cell reporter systems, SunRISER can introduce significant changes to your GOI and alter its normal function. We recommend that orthogonal techniques should be used where possible to verify the biological results obtained using SunRISER. Photobleaching of fluorescent reporters in wide-field fluorescent microscopes can vary widely between different microscope setups. In our experience with HeLa cells, 24-hour imaging with 10' intervals between frames is routinely accomplished with a DeltaVision Elite microscope. We believe with high signal-to-background, SunRISER is generally resistant to photobleaching and capable of even longer imaging experiments than described here, although this may depend on cell culture conditions and may vary between cell types. We note that we have used SunRISER successfully in human cancer cell lines and although we expect SunRISER will work in many cellular systems, extensions to primary cell lines and in vivo experiments have yet to be established. Settings for the exposure, axial spacing, and time-lapse duration may require optimization specific to other experimental designs that balance photostability, phototoxicity, as well as mRNA signal-to-background and spot detection. See (Guo and Lee, 2022) for full details on long-term imaging of mRNA molecules with SunRISER.

## Troubleshooting

### Problem 1:

Decreased cell viability after SunRISER delivery.

### Potential solution:

To reduce the toxicity associated with transfection, endotoxin-free plasmid can be used and the specific conditions of transfection (e.g., amount of FuGENE<sup>®</sup> HD Transfection Reagent/DNA mixture, incubation time) should be considered.

### Problem 2:

High fluorescent background in cells and low signal-to-background for SunRISER-labeled mRNAs.

### Potential solution:

Excessive GFP concentrations within the cells can result in a high background intensity and affect signal-to-background. Reducing the expression level of GFP could lower the basal fluorescence in transfected cells. Also check the microscope setting to find exposures that achieve the recommended intensity range. In addition, SunRISER is optimized to label mRNAs with expression levels up to several thousand transcripts per cell. GOIs expressed with higher mRNA numbers may require additional optimization. In HeLa cells, cmv and ubc promoters have been shown to achieve a 5:1 ratio of protein expression. In different target cells it may be necessary to confirm that these two promoters produce the optimal expression ratio for SunRISER components.

### Problem 3:

Highly variable intensity and size of mRNA spots.

### Potential solution:

When SunRISER is introduced to cells via transient transfection, different cells will show significant variations depending on the amount of DNA they receive. We advise choosing cells that express SunRISER at relatively low to moderate levels to focus on cells that are not undergoing stress responses to extreme overexpression. Within one cell, we expect the intensity and size of spots are reasonably consistent and should appear as diffraction-limited spots bounded by the physical limits of the microscope. Significant variation between spot intensity within the same cell could result from the suboptimal ratio of SunRISER components, and we would generally consider this cell unsuitable for imaging or further analysis. If suboptimal labeling is frequent, we recommend switching to low passage cells and check routinely for Mycoplasma.

### Problem 4:

Loss of focus during long-term imaging.

### Potential solution:

Check autofocus settings and consider reducing the number of positions imaged within a single experiment. We typically do not choose regions of the cell culture well that are confluent and instead select areas where at least 20% of the region of interest is unoccupied.

### Problem 5:

Spots are present in the image but dNEMO does not appear to be detecting all spots.

### Potential solutions:

Confirm that the axial resolution value is set appropriately. If imaging 3D stacks with a larger distance between each slice ( $> 1 \mu\text{m}$ ) it is advisable to reduce the axial resolution to ensure most diffraction-limited spots are detected. Also check settings in dNEMO to be sure spots are not being omitted for not being detected in at least the minimum number of slices (Frame Limit, Fig. 4B). The 3D image can be viewed slice-by-slice in the dNEMO interface by navigating to Image > 3D Display and selecting 'Full 3D Stack'.

Confirm that any user-defined thresholds for spot properties are not defined too narrowly so as to omit spots from consideration (Fig. 6A, B). User-defined spot feature keyframes can be examined and deleted in the 'Keyframes' panel by clicking on feature keyframes in the 'Manual Curation' dropdown menu (Fig. 6C).

### Problem 6:

Cellpose is unable to run from dNEMO (Cell Masks > Run Cellpose).

### Potential solutions:

Consult with the 'READ\_ME' file in dNEMO's documentation on <https://github.com/reaclelab> to ensure that the associated text file storing the location to your system's copy of Python3 is correct. Cellpose requires Python 3 to run, and MATLAB can by default navigate to Python 2 when both Python 2 and 3 are installed on the system.

Resulting masks from Cellpose or another automated segmentation software can also be imported into dNEMO by navigating to Cell Masks > Import Cell Mask TIFF (Fig. 5A).

### Problem 7:

The segmentations imported from Cellpose are inaccurate and need to be altered or deleted.

### Potential solutions:

Segmentations imported from Cellpose can be adjusted by either clicking on 'Modify Cell' in the 'Cells' panel to modify the given segmentation or by right-clicking the cell in question from the 'Cells' section of the 'Keyframes' panel and selecting 'Quick Seg. Redraw' to manually re-segment the cell (Fig. 5D).

The start and stop point of a tracked cell can be similarly adjusted. Right click on the cell in question in the 'Cells' section of the 'Keyframes' panel and select 'Reset Cell Start/Stop' to set the start and stop point of a given cell. This will delete any segmentations of the cell that lie outside the range of frames defined as the cell's starting and stopping time point (Fig. 5D).

### Problem 8:

Individual objects are being detected as clusters of smaller objects (oversegmentation).

### Potential solutions:

If oversegmentation of valid detected spots is broadly happening, the 'Wavelet Level' parameter may need to be increased in the detection settings. Navigate to the Settings > Signal Parameters in the interface (Fig. 4B) and increase the Wavelet Level setting to detect larger objects as single spots. Note that this will result in less reliable detection of smaller objects in the image.

There is an additional operation that is meant to reduce instances of oversegmentation in dNEMO. This operation is set on by default but may have been turned off inadvertently. Navigate to this setting in Settings > Signal Parameters and confirm that the 'Oversegmentation Check' setting is set to 'Yes'.

If the image contains artifacts which are larger than diffraction-limited spots (Fig. 2C) and are clearly not mRNA transcripts that dNEMO is identifying as multiple valid spots, the manual exclusion tool can be used to remove the artifacts from consideration (Fig. 6C, right).

## Resource availability

### **Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Robin E. C. Lee ([robinlee@pitt.edu](mailto:robinlee@pitt.edu)).

### **Materials availability**

All reagents used have been cited in the key resources table. Plasmids generated in this study are available from Addgene (plasmid IDs: 185794-185802).

### **Data and code availability**

- All data in the paper are available from Mendeley Data (DOI: 10.17632/8j4x6dj2f7.1).

- All original code been deposited on a Github repository (<https://github.com/recelelab/>), as well as on Mendeley Data where executable files are also available (DOI: 10.17632/8j4x6dj2f7.1).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## Acknowledgments

This work was supported in part to funding to R.E.C.L. by NIH grant R35-GM119462 and the Alfred P. Sloan Foundation.

## Author contributions

Conceptualization: Y.G., G.J.K., R.E.C.L. Methodology: Y.G., G.J.K., R.E.C.L. Software: G.J.K. Formal Analysis: Y.G., G.J.K., R.E.C.L. Investigation: Y.G., R.E.C.L. Writing—original draft: Y.G., G.J.K., R.E.C.L. Writing—review and editing: Y.G., G.J.K., R.E.C.L. Visualization: Y.G., G.J.K., R.E.C.L. Supervision and funding acquisition: R.E.C.L.

## Declaration of interests

The authors declare that they have no competing interests.

## References

- Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., and Long, R.M. (1998). Localization of ASH1 mRNA Particles in Living Yeast. *Molecular Cell* 2, 437-445. [https://doi.org/10.1016/S1097-2765\(00\)80143-4](https://doi.org/10.1016/S1097-2765(00)80143-4).
- Cawte, A.D., Unrau, P.J., and Rueda, D.S. (2020). Live cell imaging of single RNA molecules with fluorogenic Mango II arrays. *Nature Communications* 11, 1283. 10.1038/s41467-020-14932-7.
- Chao, J.A., Patskovsky, Y., Almo, S.C., and Singer, R.H. (2008). Structural basis for the coevolution of a viral RNA-protein complex. *Nat Struct Mol Biol* 15, 103-105. 10.1038/nsmb1327.
- Femino, A.M., Fay, F.S., Fogarty, K., and Singer, R.H. (1998). Visualization of Single RNA Transcripts in Situ. *Science* 280, 585. 10.1126/science.280.5363.585.
- Guo, Y., and Lee, R.E.C. (2022). Long-term imaging of individual mRNA molecules in living cells. *Cell Reports Methods* 2. 10.1016/j.crmeth.2022.100226.

Kowalczyk, G.J., Cruz, J.A., Guo, Y., Zhang, Q., Sauerwald, N., and Lee, R.E.C. (2021). dNEMO: a tool for quantification of mRNA and punctate structures in time-lapse images of single cells. *Bioinformatics* *37*, 677-683. [10.1093/bioinformatics/btaa874](https://doi.org/10.1093/bioinformatics/btaa874).

Linkert, M., Rueden, C.T., Allan, C., Burel, J.-M., Moore, W., Patterson, A., Loranger, B., Moore, J., Neves, C., MacDonald, D., et al. (2010). Metadata matters: access to image data in the real world. *Journal of Cell Biology* *189*, 777-782. [10.1083/jcb.201004104](https://doi.org/10.1083/jcb.201004104).

Raj, A., and van Oudenaarden, A. (2008). Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* *135*, 216-226. <https://doi.org/10.1016/j.cell.2008.09.050>.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* *9*, 671. [10.1038/nmeth.2089](https://doi.org/10.1038/nmeth.2089).

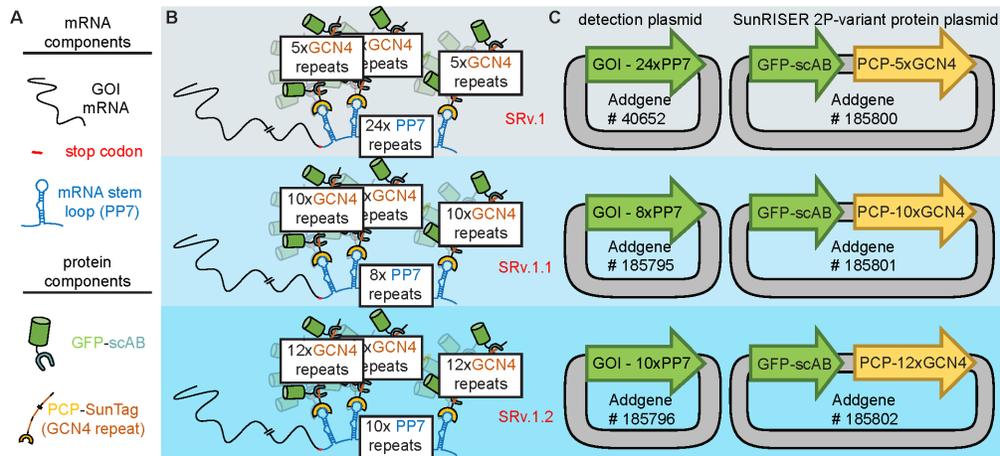
Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods* *18*, 100-106. [10.1038/s41592-020-01018-x](https://doi.org/10.1038/s41592-020-01018-x).

Tanenbaum, Marvin E., Gilbert, Luke A., Qi, Lei S., Weissman, Jonathan S., and Vale, Ronald D. (2014). A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell* *159*, 635-646. <https://doi.org/10.1016/j.cell.2014.09.039>.

Tantale, K., Mueller, F., Kozulic-Pirher, A., Lesne, A., Victor, J.-M., Robert, M.-C., Capozzi, S., Chouaib, R., Bäcker, V., Mateos-Langerak, J., et al. (2016). A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nature Communications* *7*, 12248. [10.1038/ncomms12248](https://doi.org/10.1038/ncomms12248).

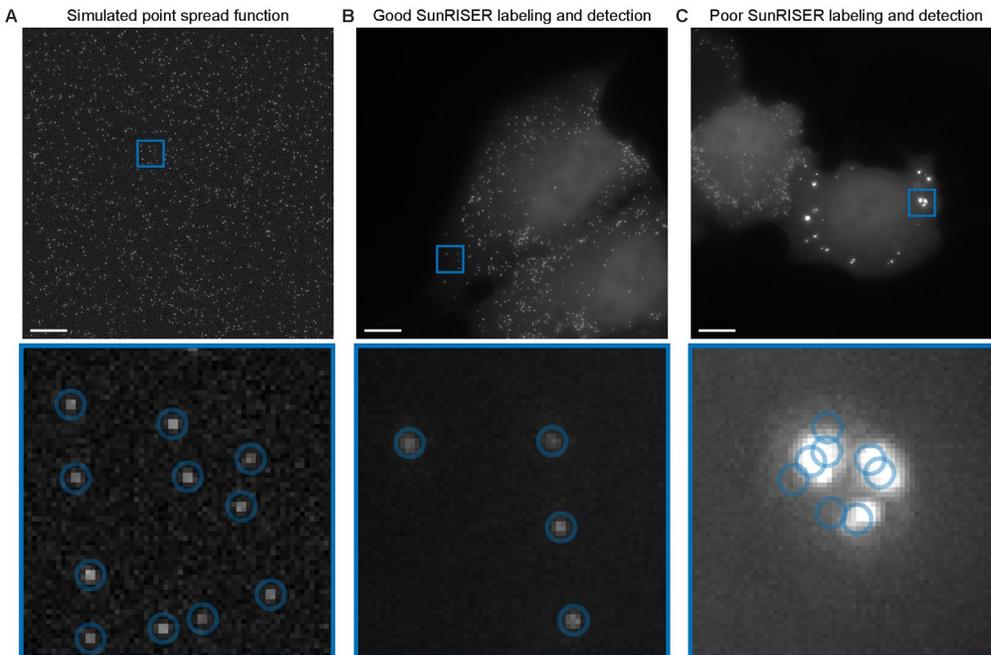
Wu, B., Chao, J.A., and Singer, R.H. (2012). Fluorescence fluctuation spectroscopy enables quantitative imaging of single mRNAs in living cells. *Biophys J* *102*, 2936-2944. [10.1016/j.bpj.2012.05.017](https://doi.org/10.1016/j.bpj.2012.05.017).

## Figures



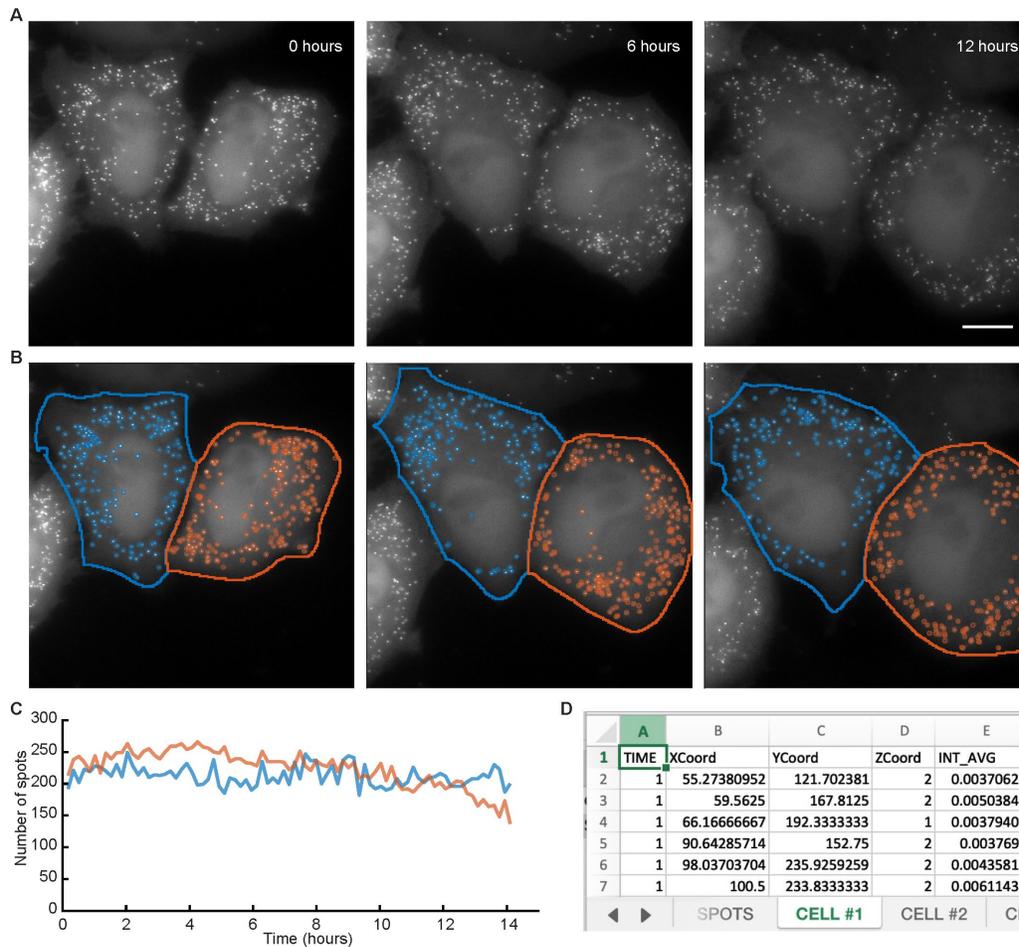
**Figure 1: Schematic of components and plasmid combinations for typical SunRISER mRNA-labeling experiments.**

**(A)** mRNA and protein components for SunRISER labeling experiments. **(B)** Schematics of SunRISER variants SRv.1 (top), SRv.1.1 (center), and SRv.1.2 (bottom) for fluorescence signal amplification. The mRNA transcript (black) is tagged at 3' UTR with PP7 stem loops (blue). In the first stage of signal amplification, each stem loop can be bound by two PCP coat proteins (yellow) fused to a SunTag GCN4 peptide array (orange). In the second stage of signal amplification, GFP (green) is recruited through antibody-peptide-specific binding between scFv (gray) and GCN4 epitopes. **(C)** Plasmid maps of the SunRISER two-plasmid (2P) variants consisting of detection plasmids (left) and protein plasmids for SRv.1-2P (top), SRv.1.1-2P (center), and SRv.1.2-2P (bottom). Using the indicated plasmids from Addgene, the GOI CDS for SRv.1 (top, left) is CFP and the GOI CDS for SRv.1.1 and SRv.1.2 (center and bottom, left) is mCherry. Although the 2P variants of SunRISER are simpler to work with, SunRISER variants using three plasmids as described previously in (Guo and Lee, 2022) can be used as outlined previously and in Table 1.

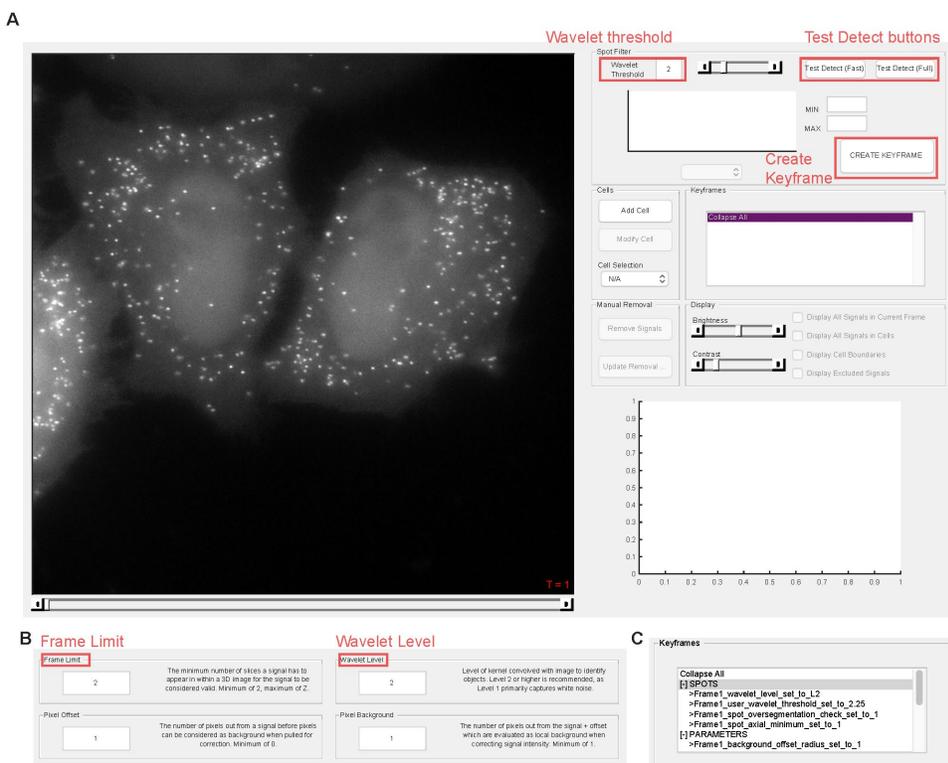


**Figure 2: SunRISER-labeled mRNAs appear as diffraction-limited spots.**

(A) Image of theoretical point spread functions (PSFs) for diffraction limited signals. The simulated image was generated as described previously (see (Kowalczyk *et al.*, 2021)). (B, C) Representative maximum intensity projections of HeLa cells transfected with SunRISER SRv.1-2P and detection plasmid CFP-24xPP7 (top images). Bottom images represent detail of fluorescence images as indicated, with blue circles representing spots detected by dNEMO analysis. In good labeling conditions, (B) mRNA molecules labeled with SunRISER appear as diffraction-limited spots comparable to those generated in the simulated image. In poor labeling (C) fluorescent structures are larger than diffraction-limited objects and do not represent single mRNA molecules. Large objects also show evidence of oversegmentation where a single fluorescent structure is detected as multiple spots.

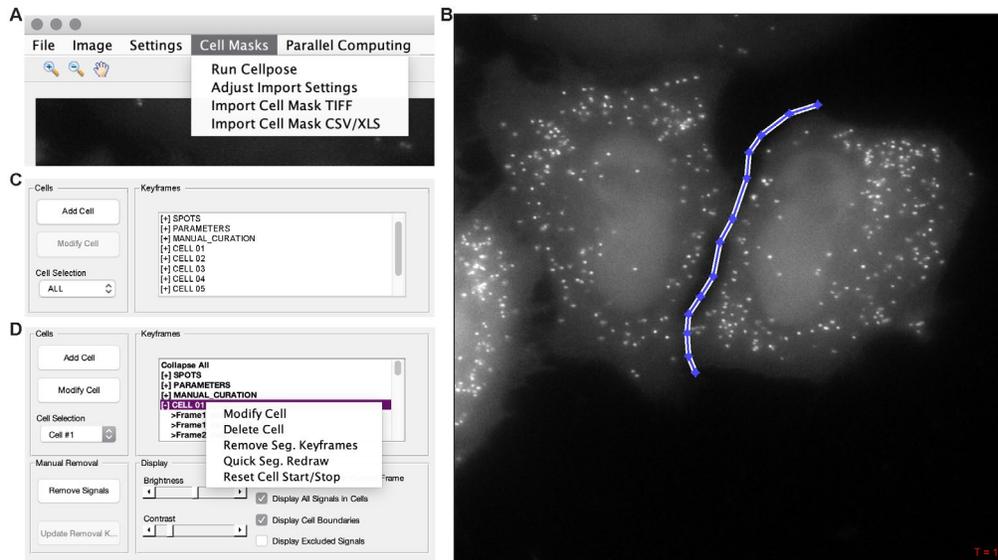


**Figure 3: Quantification of SunRISER-labeled mRNAs in time-lapse live-cell images with dNEMO.** **(A)** Maximum intensity projection of HeLa cells transfected with SunRISER SRv.1-2P and detection plasmid CFP-24xPP7. Cells were imaged every 10 minutes for 24 hours. **(B)** SunRISER-labeled mRNAs detected by dNEMO (blue or orange circles) and associated with cells segmented using Cellpose. **(C)** Time-courses for the number of mRNA molecules identified within the 2 cells shown in (A). **(D)** Example results file generated by dNEMO and output to excel. Features are collected for every spot, and spots within each single cell are separated by tabs.



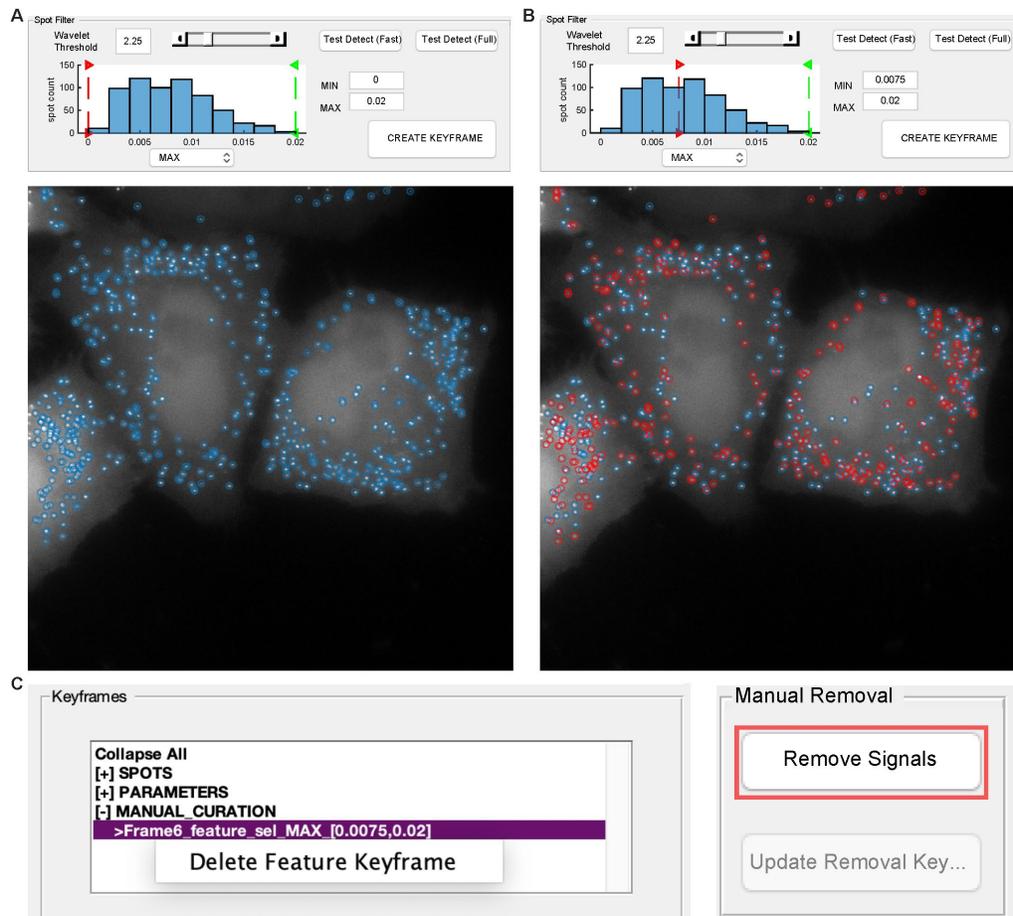
**Figure 4: Spot detection settings and operations in dNEMO.**

**(A)** The dNEMO software interface with an open image of HeLa cells transfected with SunRISER SRv.1-2P and detection plasmid CFP-24xPP7. Highlighted within the ‘Spot Filter’ panel (upper-right) are critical user operations: the ‘Wavelet Threshold’ value; the detection of spots over the currently displayed image (‘Test Detect’ buttons); and the generation of a keyframe of the current detection settings (‘Create Keyframe’ button). **(B)** Settings GUI accessible in dNEMO which holds user-defined parameters for spot detection. Highlighted are the ‘Frame Limit’ parameter (upper left) and the ‘Wavelet Level’ parameter (upper right). **(C)** Screenshot of the ‘Keyframes’ panel after the spot detection operation has completed operating over the time-lapse images.



**Figure 5: dNEMO interfaces with Cellpose for automated segmentation of cells.**

**(A)** Screenshot of ‘Cell Masks’ drop-down menu to access Cellpose for cell segmentation over the currently displayed image as well as importing previously generated cell segmentation mask files into dNEMO. **(B)** Manual segmentation function operating over a displayed image in dNEMO with a user-defined polygon. **(C)** Screenshot of the ‘Cells’ and ‘Keyframes’ panels after Cellpose has completed segmentation and imported the resulting masks into dNEMO. **(D)** Screenshot of the user operations available to edit imported Cellpose results within the dNEMO interface after right clicking a cell within the ‘Keyframes’ panel.



**Figure 6: Keyframing and manual exclusion tools for curation of spot datasets.**

(A, B) Screenshots of the feature selection in dNEMO for user-assisted filtering of spots based on their size or intensity, among other features. User-defined parameters for the maximum intensity of detected spots set with a lower bound of 0 (A, top) or 0.0075 (B, top). Resulting dNEMO spots either accepted (blue) or filtered (red) based on analysis with the values as shown (bottom). (C) Keyframe information window after clicking the ‘Create Keyframe’ button using the intensity-filtering settings in B (Left). The results of the filter can be deleted by clicking the entry. The ‘Manual Removal’ panel (Right) with the spot removal operation indicated (highlighted, red) is used to supplement user-assisted filtering, enabling keyframe entries for removal of user-selected spots from further analysis.

## Bibliography

1. Huang, S., *Non-genetic heterogeneity of cells in development: more than just noise*. Development, 2009. **136**(23): p. 3853-62.
2. Altschuler, S.J. and L.F. Wu, *Cellular heterogeneity: do differences make a difference?* Cell, 2010. **141**(4): p. 559-63.
3. Brock, A., H. Chang, and S. Huang, *Non-genetic heterogeneity — a mutation-independent driving force for the somatic evolution of tumours*. Nature Reviews Genetics, 2009. **10**(5): p. 336-342.
4. Huh, D. and J. Paulsson, *Non-genetic heterogeneity from stochastic partitioning at cell division*. Nature Genetics, 2011. **43**(2): p. 95-100.
5. Goldman, S.L., et al., *The Impact of Heterogeneity on Single-Cell Sequencing*. Front Genet, 2019. **10**: p. 8.
6. Singh, A. and M. Soltani, *Quantifying intrinsic and extrinsic variability in stochastic gene expression models*. PLoS One, 2013. **8**(12): p. e84301.
7. Elowitz, M.B., et al., *Stochastic gene expression in a single cell*. Science, 2002. **297**(5584): p. 1183-6.
8. Scott, M., B. Ingalls, and M. Kaern, *Estimations of intrinsic and extrinsic noise in models of nonlinear genetic networks*. Chaos, 2006. **16**(2): p. 026107.
9. Hilfinger, A. and J. Paulsson, *Separating intrinsic from extrinsic fluctuations in dynamic biological systems*. Proc Natl Acad Sci U S A, 2011. **108**(29): p. 12167-72.
10. Yosef, N. and A. Regev, *Impulse control: temporal dynamics in gene transcription*. Cell, 2011. **144**(6): p. 886-96.
11. Pascual-Ahuir, A., J. Fita-Torró, and M. Proft, *Capturing and Understanding the Dynamics and Heterogeneity of Gene Expression in the Living Cell*. Int J Mol Sci, 2020. **21**(21).
12. Lenstra, T.L., et al., *Transcription Dynamics in Living Cells*. Annual review of biophysics, 2016. **45**: p. 25-47.
13. Coulon, A., et al., *Eukaryotic transcriptional dynamics: from single molecules to cell populations*. Nat Rev Genet, 2013. **14**(8): p. 572-84.
14. Kholodenko, B.N., *Cell-signalling dynamics in time and space*. Nature Reviews Molecular Cell Biology, 2006. **7**(3): p. 165-176.

15. Hoffmann, A., et al., *The IkappaB-NF-kappaB signaling module: temporal control and selective gene activation*. Science, 2002. **298**(5596): p. 1241-5.
16. Murphy, L.O., et al., *Molecular interpretation of ERK signal duration by immediate early gene products*. Nat Cell Biol, 2002. **4**(8): p. 556-64.
17. Levsky, J.M., et al., *Single-cell gene expression profiling*. Science, 2002. **297**(5582): p. 836-40.
18. Larson, D.R., R.H. Singer, and D. Zenklusen, *A single molecule view of gene expression*. Trends Cell Biol, 2009. **19**(11): p. 630-7.
19. Vera, M., et al., *Single-Cell and Single-Molecule Analysis of Gene Expression Regulation*. Annu Rev Genet, 2016. **50**: p. 267-291.
20. Patange, S., M. Girvan, and D.R. Larson, *Single-cell systems biology: probing the basic unit of information flow*. Curr Opin Syst Biol, 2018. **8**: p. 7-15.
21. Yu, T. and J. Scolnick, *Complex biological questions being addressed using single cell sequencing technologies*. SLAS Technology, 2022. **27**(2): p. 143-149.
22. Qian, Z. and L. Bao, *Chapter 3 - Cellular Heterogeneity and Single-Cell Omics*, in *Single-Cell Omics*, D. Barh and V. Azevedo, Editors. 2019, Academic Press. p. 35-44.
23. Lee, J., D.Y. Hyeon, and D. Hwang, *Single-cell multiomics: technologies and data analysis methods*. Experimental & Molecular Medicine, 2020. **52**(9): p. 1428-1442.
24. Anselmetti, D., *Single cell analysis: technologies and applications*. 2009: John Wiley & Sons.
25. McKinnon, K.M., *Flow Cytometry: An Overview*. Curr Protoc Immunol, 2018. **120**: p. 5.1.1-5.1.11.
26. Schwartzman, O. and A. Tanay, *Single-cell epigenomics: techniques and emerging applications*. Nature Reviews Genetics, 2015. **16**(12): p. 716-726.
27. Gawad, C., W. Koh, and S.R. Quake, *Single-cell genome sequencing: current state of the science*. Nature Reviews Genetics, 2016. **17**(3): p. 175-188.
28. Prakadan, S.M., A.K. Shalek, and D.A. Weitz, *Scaling by shrinking: empowering single-cell'omics' with microfluidic devices*. Nature Reviews Genetics, 2017. **18**(6): p. 345-361.
29. Tang, X., et al., *The single-cell sequencing: new developments and medical applications*. Cell & Bioscience, 2019. **9**(1): p. 53.
30. Kashima, Y., et al., *Single-cell sequencing techniques from individual to multiomics analyses*. Experimental & Molecular Medicine, 2020. **52**(9): p. 1419-1427.

31. Hong, T.H. and W.-Y. Park, *Single-cell genomics technology: perspectives*. Experimental & Molecular Medicine, 2020. **52**(9): p. 1407-1408.
32. Aldridge, S. and S.A. Teichmann, *Single cell transcriptomics comes of age*. Nature Communications, 2020. **11**(1): p. 4307.
33. Slavov, N., *Unpicking the proteome in single cells*. Science, 2020. **367**(6477): p. 512-513.
34. Alfaro, J.A., et al., *The emerging landscape of single-molecule protein sequencing technologies*. Nature Methods, 2021. **18**(6): p. 604-617.
35. Zhang, Y., et al., *Single-cell RNA sequencing in cancer research*. Journal of Experimental & Clinical Cancer Research, 2021. **40**(1): p. 81.
36. Teichmann, S. and A. Regev, *The network effect: studying COVID-19 pathology with the Human Cell Atlas*. Nat Rev Mol Cell Biol, 2020. **21**(8): p. 415-416.
37. Tyson, J.J., K.C. Chen, and B. Novak, *Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell*. Curr Opin Cell Biol, 2003. **15**(2): p. 221-31.
38. Purvis, J.E. and G. Lahav, *Encoding and decoding cellular information through signaling dynamics*. Cell, 2013. **152**(5): p. 945-56.
39. Nandagopal, N., et al., *Dynamic ligand discrimination in the notch signaling pathway*. Cell, 2018. **172**(4): p. 869-880. e19.
40. Nelson, D., et al., *Oscillations in NF- $\kappa$ B signaling control the dynamics of gene expression*. Science, 2004. **306**(5696): p. 704-708.
41. Covert, M.W., et al., *Achieving stability of lipopolysaccharide-induced NF- $\kappa$ B activation*. Science, 2005. **309**(5742): p. 1854-1857.
42. Batchelor, E., et al., *Stimulus-dependent dynamics of p53 in single cells*. Molecular systems biology, 2011. **7**(1): p. 488.
43. Purvis, J.E., et al., *p53 dynamics control cell fate*. Science, 2012. **336**(6087): p. 1440-4.
44. Weake, V.M. and J.L. Workman, *Inducible gene expression: diverse regulatory mechanisms*. Nat Rev Genet, 2010. **11**(6): p. 426-37.
45. de Nadal, E., G. Ammerer, and F. Posas, *Controlling gene expression in response to stress*. Nat Rev Genet, 2011. **12**(12): p. 833-45.
46. Pérez-Ortín, J.E., et al., *Eukaryotic mRNA decay: methodologies, pathways, and links to other stages of gene expression*. J Mol Biol, 2013. **425**(20): p. 3750-75.
47. Vihervaara, A., F.M. Duarte, and J.T. Lis, *Molecular mechanisms driving transcriptional stress responses*. Nat Rev Genet, 2018. **19**(6): p. 385-397.

48. Moore Melissa, J., *From Birth to Death: The Complex Lives of Eukaryotic mRNAs*. Science, 2005. **309**(5740): p. 1514-1518.
49. Hocine, S., R.H. Singer, and D. Grünwald, *RNA processing and export*. Cold Spring Harb Perspect Biol, 2010. **2**(12): p. a000752.
50. Schoenberg, D.R. and L.E. Maquat, *Regulation of cytoplasmic mRNA decay*. Nature Reviews Genetics, 2012. **13**(4): p. 246-259.
51. Crick, F., *Central dogma of molecular biology*. Nature, 1970. **227**(5258): p. 561-563.
52. Shapiro, J.A., *Revisiting the Central Dogma in the 21st Century*. Annals of the New York Academy of Sciences, 2009. **1178**(1): p. 6-28.
53. Koonin, E.V., *Does the central dogma still stand?* Biology Direct, 2012. **7**(1): p. 27.
54. Ille, A.M., H. Lamont, and M.B. Mathews, *The Central Dogma revisited: Insights from protein synthesis, CRISPR, and beyond*. WIREs RNA, 2022. **n/a**(n/a): p. e1718.
55. Baltimore, D., *Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses*. Nature, 1970. **226**(5252): p. 1209-1211.
56. Temin, H.M. and S. Mizutani, *Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of Rous Sarcoma Virus*. Nature, 1970. **226**(5252): p. 1211-1213.
57. Roberts, J.D., K. Bebenek, and T.A. Kunkel, *The Accuracy of Reverse Transcriptase from HIV-1*. Science, 1988. **242**(4882): p. 1171-1173.
58. González-Delgado, A., et al., *Prokaryotic reverse transcriptases: from retroelements to specialized defense systems*. FEMS Microbiology Reviews, 2021. **45**(6): p. fuab025.
59. Temin, H.M., *Reverse transcription in the eukaryotic genome: retroviruses, pararetroviruses, retrotransposons, and retrotranscripts*. Mol Biol Evol, 1985. **2**(6): p. 455-68.
60. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics*. Nat Rev Genet, 2009. **10**(1): p. 57-63.
61. Alonso, D. and A. Mondragón, *Mechanisms of catalytic RNA molecules*. Biochemical Society Transactions, 2021. **49**(4): p. 1529-1535.
62. Fejes-Toth, K., et al., *Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs*. Nature, 2009. **457**(7232): p. 1028-1032.
63. McNally, L., U. Manne, and W.E. Grizzle, *Post-transcriptional processing of genetic information and its relation to cancer*. Biotech Histochem, 2013. **88**(7): p. 365-72.

64. Deribe, Y.L., T. Pawson, and I. Dikic, *Post-translational modifications in signal integration*. Nature Structural & Molecular Biology, 2010. **17**(6): p. 666-672.
65. Jiricny, J., *The multifaceted mismatch-repair system*. Nat Rev Mol Cell Biol, 2006. **7**(5): p. 335-46.
66. Goodman, M.F., *Error-prone repair DNA polymerases in prokaryotes and eukaryotes*. Annu Rev Biochem, 2002. **71**: p. 17-50.
67. Bandyra, K. and B. Luisi, *Central dogma alchemy*. Rna, 2015. **21**(4): p. 558-9.
68. Baudrimont, A., et al., *Contribution of RNA Degradation to Intrinsic and Extrinsic Noise in Gene Expression*. Cell Rep, 2019. **26**(13): p. 3752-3761.e5.
69. Tuck, A.C., et al., *Mammalian RNA Decay Pathways Are Highly Specialized and Widely Linked to Translation*. Mol Cell, 2020. **77**(6): p. 1222-1236.e13.
70. Kloc, M., N.R. Zearfoss, and L.D. Etkin, *Mechanisms of Subcellular mRNA Localization*. Cell, 2002. **108**(4): p. 533-544.
71. Holt, C.E. and S.L. Bullock, *Subcellular mRNA localization in animal cells and why it matters*. Science, 2009. **326**(5957): p. 1212-6.
72. Akam, M.E., *The location of Ultrabithorax transcripts in Drosophila tissue sections*. The EMBO journal, 1983. **2**(11): p. 2075-2084.
73. Gaspar, I. and A. Ephrussi, *Strength in numbers: quantitative single-molecule RNA detection assays*. WIREs Developmental Biology, 2015. **4**(2): p. 135-150.
74. Singer, R.H. and D.C. Ward, *Actin gene expression visualized in chicken muscle tissue culture by using in situ hybridization with a biotinylated nucleotide analog*. Proceedings of the National Academy of Sciences, 1982. **79**(23): p. 7331-7335.
75. Femino, A.M., et al., *Visualization of single RNA transcripts in situ*. Science, 1998. **280**(5363): p. 585-590.
76. Raj, A., et al., *Imaging individual mRNA molecules using multiple singly labeled probes*. Nature Methods, 2008. **5**(10): p. 877-879.
77. Lee, R.E., et al., *Fold change of nuclear NF- $\kappa$ B determines TNF-induced transcription in single cells*. Molecular cell, 2014. **53**(6): p. 867-879.
78. Chao, J.A. and T. Lionnet, *Imaging the Life and Death of mRNAs in Single Cells*. Cold Spring Harb Perspect Biol, 2018. **10**(12).
79. Bertrand, E., et al., *Localization of ASH1 mRNA Particles in Living Yeast*. Molecular Cell, 1998. **2**(4): p. 437-445.

80. Buxbaum, A.R., B. Wu, and R.H. Singer, *Single  $\beta$ -actin mRNA detection in neurons reveals a mechanism for regulating its translatability*. *Science* (New York, N.Y.), 2014. **343**(6169): p. 419-422.
81. Czaplinski, K., *Techniques for Single-Molecule mRNA Imaging in Living Cells*, in *Neuroepigenomics in Aging and Disease*, R. Delgado-Morales, Editor. 2017, Springer International Publishing: Cham. p. 425-441.
82. Horvathova, I., et al., *The Dynamics of mRNA Turnover Revealed by Single-Molecule Imaging in Single Cells*. *Mol Cell*, 2017. **68**(3): p. 615-625.e9.
83. Katz, Z.B., et al., *Mapping translation 'hot-spots' in live cells by tracking single molecules of mRNA and ribosomes*. *eLife*, 2016. **5**: p. e10415.
84. Khuperkar, D., et al., *Quantification of mRNA translation in live cells using single-molecule imaging*. *Nature Protocols*, 2020. **15**(4): p. 1371-1398.
85. Larson, D.R., et al., *Real-time observation of transcription initiation and elongation on an endogenous yeast gene*. *Science* (New York, N.Y.), 2011. **332**(6028): p. 475-478.
86. Mateju, D., et al., *Single-Molecule Imaging Reveals Translation of mRNAs Localized to Stress Granules*. *Cell*, 2020. **183**(7): p. 1801-1812.e13.
87. Moon, S.L., et al., *Multicolour single-molecule tracking of mRNA interactions with RNP granules*. *Nature Cell Biology*, 2019. **21**(2): p. 162-168.
88. Mor, A., et al., *Dynamics of single mRNP nucleocytoplasmic transport and export through the nuclear pore in living cells*. *Nature Cell Biology*, 2010. **12**(6): p. 543-552.
89. Sato, H., et al., *Imaging of DNA and RNA in Living Eukaryotic Cells to Reveal Spatiotemporal Dynamics of Gene Expression*. *Annual Review of Biochemistry*, 2020. **89**(1): p. 159-187.
90. Vargas, D.Y., et al., *Single-molecule imaging of transcriptionally coupled and uncoupled splicing*. *Cell*, 2011. **147**(5): p. 1054-65.
91. Vitor, A.C., et al., *Single-molecule imaging of transcription at damaged chromatin*. *Science Advances*, 2019. **5**(1): p. eaau1249.
92. Wan, Y., et al., *Dynamic imaging of nascent RNA reveals general principles of transcription dynamics and stochastic splice site selection*. *Cell*, 2021. **184**(11): p. 2878-2895.e20.
93. Wang, C., et al., *Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells*. *Cell*, 2016. **165**(4): p. 990-1001.
94. Wilbertz, J.H., et al., *Single-Molecule Imaging of mRNA Localization and Regulation during the Integrated Stress Response*. *Molecular Cell*, 2019. **73**(5): p. 946-958.e7.

95. Wu, B., et al., *Translation dynamics of single mRNAs in live cells and neurons*. Science, 2016. **352**(6292): p. 1430-1435.
96. Yan, X., et al., *Dynamics of Translation of Single mRNA Molecules In Vivo*. Cell, 2016. **165**(4): p. 976-989.
97. Pichon, X., et al., *A Growing Toolbox to Image Gene Expression in Single Cells: Sensitive Approaches for Demanding Challenges*. Molecular Cell, 2018. **71**(3): p. 468-480.
98. Tutucci, E., et al., *Imaging mRNA In Vivo, from Birth to Death*. Annual Review of Biophysics, 2018. **47**(1): p. 85-106.
99. Braselmann, E., et al., *Illuminating RNA Biology: Tools for Imaging RNA in Live Mammalian Cells*. Cell Chemical Biology, 2020. **27**(8): p. 891-903.
100. Fusco, D., et al., *Single mRNA molecules demonstrate probabilistic movement in living mammalian cells*. Current biology : CB, 2003. **13**(2): p. 161-167.
101. Rath, A.K. and A. Rentmeister, *Genetically encoded tools for RNA imaging in living cells*. Current Opinion in Biotechnology, 2015. **31**: p. 42-49.
102. Vera, M., et al., *Single-Cell and Single-Molecule Analysis of Gene Expression Regulation*. Annual review of genetics, 2016. **50**: p. 267-291.
103. Ben-Ari, Y., et al., *The life of an mRNA in space and time*. Journal of Cell Science, 2010. **123**(10): p. 1761.
104. Ferguson, M.L. and D.R. Larson, *Measuring Transcription Dynamics in Living Cells Using Fluctuation Analysis*, in *Imaging Gene Expression: Methods and Protocols*, Y. Shav-Tal, Editor. 2013, Humana Press: Totowa, NJ. p. 47-60.
105. Lenstra, T.L. and D.R. Larson, *Single-Molecule mRNA Detection in Live Yeast*. Current protocols in molecular biology, 2016. **113**: p. 14.24.1-14.24.15.
106. Spille, J. and U. Kubitscheck, *Labelling and imaging of single endogenous messenger RNA particles in vivo*. Journal of Cell Science, 2015. **128**: p. 3695 - 3706.
107. Tutucci, E., et al., *An improved MS2 system for accurate reporting of the mRNA life cycle*. Nature methods, 2018. **15**(1): p. 81-89.
108. Wu, B., J.A. Chao, and R.H. Singer, *Fluorescence fluctuation spectroscopy enables quantitative imaging of single mRNAs in living cells*. Biophysical journal, 2012. **102**(12): p. 2936-2944.
109. Forero-Quintero, L.S., et al., *Live-cell imaging reveals the spatiotemporal organization of endogenous RNA polymerase II phosphorylation at a single gene*. Nature Communications, 2021. **12**(1): p. 3158.

110. Golding, I. and E.C. Cox, *RNA dynamics in live Escherichia coli cells*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(31): p. 11310-11315.
111. Tantale, K., et al., *A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting*. Nature Communications, 2016. **7**(1): p. 12248.
112. Brewster, Robert C., et al., *The Transcription Factor Titration Effect Dictates Level of Gene Expression*. Cell, 2014. **156**(6): p. 1312-1323.
113. Kosugi, S., et al., *Six Classes of Nuclear Localization Signals Specific to Different Binding Grooves of Importin  $\alpha^*$* . Journal of Biological Chemistry, 2009. **284**(1): p. 478-485.
114. Róna, G., et al., *NLS copy-number variation governs efficiency of nuclear import – case study on dUTPases*. The FEBS Journal, 2014. **281**(24): p. 5463-5478.
115. Zhang, Z.C., et al., *Evolutionary development of redundant nuclear localization signals in the mRNA export factor NXF1*. Molecular Biology of the Cell, 2011. **22**(23): p. 4657-4668.
116. Yao, Q., et al., *Nuclear localization of Hif-3 $\alpha$  requires two redundant NLS motifs in its unique C-terminal region*. FEBS Letters, 2018. **592**(16): p. 2769-2775.
117. Rahman, S. and D. Zenklusen, *Single-Molecule Resolution Fluorescent In Situ Hybridization (smFISH) in the Yeast S. cerevisiae*, in *Imaging Gene Expression: Methods and Protocols*, Y. Shav-Tal, Editor. 2013, Humana Press: Totowa, NJ. p. 33-46.
118. Tanenbaum, Marvin E., et al., *A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging*. Cell, 2014. **159**(3): p. 635-646.
119. Boersma, S., et al., *Multi-Color Single-Molecule Imaging Uncovers Extensive Heterogeneity in mRNA Decoding*. Cell, 2019. **178**(2): p. 458-472.e19.
120. Wickens, M., et al., *In Translational Control of Gene Expression; Sonenberg, N., Hershey, JWB, Mathews, MB, Eds.* 2000, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York.
121. Kuersten, S. and E.B. Goodwin, *The power of the 3' UTR: translational control and development*. Nature Reviews Genetics, 2003. **4**(8): p. 626-637.
122. Gebauer, F. and M.W. Hentze, *Molecular mechanisms of translational control*. Nature Reviews Molecular Cell Biology, 2004. **5**(10): p. 827-835.
123. Sonenberg, N. and A.G. Hinnebusch, *Regulation of translation initiation in eukaryotes: mechanisms and biological targets*. Cell, 2009. **136**(4): p. 731-45.
124. Hershey, J.W., N. Sonenberg, and M.B. Mathews, *Principles of translational control: an overview*. Cold Spring Harb Perspect Biol, 2012. **4**(12).

125. Gingras, A.-C., B. Raught, and N. Sonenberg, *eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation*. Annual review of biochemistry, 1999. **68**(1): p. 913-963.
126. Stebbins-Boaz, B., et al., *Maskin is a CPEB-associated factor that transiently interacts with eIF-4E*. Molecular cell, 1999. **4**(6): p. 1017-1027.
127. Mazumder, B., et al., *Regulated release of L13a from the 60S ribosomal subunit as a mechanism of transcript-specific translational control*. Cell, 2003. **115**(2): p. 187-98.
128. Capece, M.C., et al., *A simple real-time assay for in vitro translation*. Rna, 2015. **21**(2): p. 296-305.
129. Ingolia, N.T., et al., *Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling*. science, 2009. **324**(5924): p. 218-223.
130. Larsson, O., B. Tian, and N. Sonenberg, *Toward a genome-wide landscape of translational control*. Cold Spring Harb Perspect Biol, 2013. **5**(1): p. a012302.
131. Brar, G.A. and J.S. Weissman, *Ribosome profiling reveals the what, when, where and how of protein synthesis*. Nature Reviews Molecular Cell Biology, 2015. **16**(11): p. 651-664.
132. Aakalu, G., et al., *Dynamic visualization of local protein synthesis in hippocampal neurons*. Neuron, 2001. **30**(2): p. 489-502.
133. Dieterich, D.C., et al., *In situ visualization and dynamics of newly synthesized proteins in rat hippocampal neurons*. Nature neuroscience, 2010. **13**(7): p. 897-905.
134. Morisaki, T., et al., *Real-time quantification of single RNA translation dynamics in living cells*. Science, 2016. **352**(6292): p. 1425-1429.
135. Wu, B., et al., *Translation dynamics of single mRNAs in live cells and neurons*. Science (New York, N.Y.), 2016. **352**(6292): p. 1430-1435.
136. Pichon, X., et al., *Visualization of single endogenous polysomes reveals the dynamics of translation in live human cells*. Journal of Cell Biology, 2016. **214**(6): p. 769-781.
137. Akansu, A.N., W.A. Serdijn, and I.W. Selesnick, *Emerging applications of wavelets: A review*. Physical communication, 2010. **3**(1): p. 1-18.
138. Demirel, H. and G. Anbarjafari, *Image resolution enhancement by using discrete and stationary wavelet decomposition*. IEEE transactions on image processing, 2010. **20**(5): p. 1458-1460.
139. Izeddin, I., et al., *Wavelet analysis for single molecule localization microscopy*. Optics express, 2012. **20**(3): p. 2081-2095.

140. Olivo-Marin, J.-C., *Extraction of spots in biological images using multiscale products*. Pattern recognition, 2002. **35**(9): p. 1989-1996.
141. Kowalczyk, G.J., et al., *dNEMO: a tool for quantification of mRNA and punctate structures in time-lapse images of single cells*. Bioinformatics, 2021. **37**(5): p. 677–683.
142. Tulpule, A., et al., *Kinase-mediated RAS signaling via membraneless cytoplasmic protein granules*. Cell, 2021. **184**(10): p. 2649-2664.e18.
143. Pabon, N.A., et al., *A network-centric approach to drugging TNF-induced NF- $\kappa$ B signaling*. Nature communications, 2019. **10**(1): p. 1-9.
144. Cruz, J.A., et al., *A variable-gain stochastic pooling motif mediates information transfer from receptor assemblies into NF- $\kappa$ B*. Science Advances. **7**(30): p. eabi9410.
145. Du, M., et al., *Liquid phase separation of NEMO induced by polyubiquitin chains activates NF- $\kappa$ B*. Molecular Cell, 2022. **82**(13): p. 2415-2426.e5.
146. Morfill, J., et al., *Affinity-matured recombinant antibody fragments analyzed by single-molecule force spectroscopy*. Biophysical journal, 2007. **93**(10): p. 3583-3590.
147. Weber-Bornhauser, S., et al., *Thermodynamics and Kinetics of the Reaction of a Single-Chain Antibody Fragment (scFv) with the Leucine Zipper Domain of Transcription Factor GCN4*. Biochemistry, 1998. **37**(37): p. 13011-13020.
148. Voigt, F., et al., *Single-Molecule Quantification of Translation-Dependent Association of mRNAs with the Endoplasmic Reticulum*. Cell Reports, 2017. **21**(13): p. 3740-3753.
149. Guo, Y. and R.E.C. Lee, *Long-term imaging of individual mRNA molecules in living cells*. Cell Reports Methods, 2022. **2**(6): p. 100226.
150. Wörn, A., et al., *Correlation between in vitro stability and in vivo performance of anti-GCN4 intrabodies as cytoplasmic inhibitors*. The Journal of biological chemistry, 2000. **275**(4): p. 2795-2803.
151. Lim, F. and D.S. Peabody, *RNA recognition site of PP7 coat protein*. Nucleic acids research, 2002. **30**(19): p. 4138-4144.
152. Qin, J.Y., et al., *Systematic comparison of constitutive promoters and the doxycycline-inducible promoter*. PloS one, 2010. **5**(5): p. e10611-e10611.
153. Heinrich, S., et al., *Stem-loop RNA labeling can affect nuclear and cytoplasmic mRNA processing*. RNA (New York, N.Y.), 2017. **23**(2): p. 134-141.
154. Kahana, C., G. Asher, and Y. Shaul, *Mechanisms of protein degradation: an odyssey with ODC*. Cell Cycle, 2005. **4**(11): p. 1461-4.
155. AAT Bioquest, I., *"Quest Calculate™ RNA Molecular Weight Calculator."* 2022.

156. Stothard, P., *The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences*. BioTechniques, 2000. **28**(6): p. 1102-1104.
157. Katz, N., et al., *Overcoming the design, build, test bottleneck for synthesis of nonrepetitive protein-RNA cassettes*. Nature communications, 2021. **12**(1): p. 1576-1576.
158. Ma, H., et al., *CRISPR-Sirius: RNA scaffolds for signal amplification in genome imaging*. Nature methods, 2018. **15**(11): p. 928-931.
159. Saroufim, M.-A., et al., *The nuclear basket mediates perinuclear mRNA scanning in budding yeast*. Journal of Cell Biology, 2015. **211**(6): p. 1131-1140.
160. Lionnet, T., et al., *A transgenic mouse for in vivo detection of endogenous labeled mRNA*. Nat Methods, 2011. **8**(2): p. 165-70.
161. Hocine, S., et al., *Single-molecule analysis of gene expression using two-color RNA labeling in live yeast*. Nat Methods, 2013. **10**(2): p. 119-21.
162. Bahar Halpern, K., et al., *Nuclear Retention of mRNA in Mammalian Tissues*. Cell reports, 2015. **13**(12): p. 2653-2662.
163. Schwanhäusser, B., et al., *Global quantification of mammalian gene expression control*. Nature, 2011. **473**(7347): p. 337-342.
164. Tani, H., et al., *Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals*. Genome research, 2012. **22**(5): p. 947-956.
165. Jaqaman, K., et al., *Robust single-particle tracking in live-cell time-lapse sequences*. Nature Methods, 2008. **5**(8): p. 695-702.
166. Huh, D. and J. Paulsson, *Random partitioning of molecules at cell division*. Proceedings of the National Academy of Sciences, 2011. **108**(36): p. 15004.
167. Shlyakhtina, Y., K.L. Moran, and M.M. Portal, *Asymmetric Inheritance of Cell Fate Determinants: Focus on RNA*. Non-Coding RNA, 2019. **5**(2).
168. Stahl, T., et al., *Asymmetric distribution of glucose transporter mRNA provides a growth advantage in yeast*. The EMBO Journal, 2019. **38**(10): p. e100373.
169. Skamagki, M., et al., *Asymmetric Localization of Cdx2 mRNA during the First Cell-Fate Decision in Early Mouse Development*. Cell Reports, 2013. **3**(2): p. 442-457.
170. Varela, N., et al., *Mitotic Inheritance of mRNA Facilitates Translational Activation of the Osteogenic-Lineage Commitment Factor Runx2 in Progeny of Osteoblastic Cells*. Journal of cellular physiology, 2016. **231**(5): p. 1001-1014.
171. Sunchu, B. and C. Cabernard, *Principles and mechanisms of asymmetric cell division*. Development, 2020. **147**(13): p. dev167650.

172. Raj, A. and S. Tyagi, *Chapter 17 - Detection of Individual Endogenous RNA Transcripts In Situ Using Multiple Singly Labeled Probes*, in *Methods in Enzymology*, N.G. Walter, Editor. 2010, Academic Press. p. 365-386.
173. Djebali, S., et al., *Landscape of transcription in human cells*. *Nature*, 2012. **489**(7414): p. 101-108.
174. O'Brien, J., et al., *Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation*. *Frontiers in Endocrinology*, 2018. **9**.
175. Kristensen, L.S., et al., *The biogenesis, biology and characterization of circular RNAs*. *Nature Reviews Genetics*, 2019. **20**(11): p. 675-691.